

Untersuchungen zum Dotieren von Silicium aus einer
Oberflächenbelegung mit Phosphor in einem
Kurzzeitprozess

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur

(Dr.-Ing.)

von Dipl.-Phys. Bodo Kalkofen

geb. am 27.05.1970 in Stendal

genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Edmund P. Burte
Prof. Dr. Udo Schwalke
Dr. Zsolt Nènyei

Promotionskolloquium am 16.04.2007

Inhaltsverzeichnis

1.	Einleitung	5
1.1	Motivation und Zielsetzung der Untersuchungen zur Siliciumdotierung	5
1.1.1	Einordnung in die Siliciumtechnologie	5
1.1.2	Arbeitsschwerpunkte	6
1.2	Gliederung der Arbeit	7
2.	Dotierung von Silicium zur Erzeugung flacher pn-Übergänge	9
2.1	Die Dotierung im Laufe der Entwicklung der Halbleitertechnologie	9
2.2	Bedeutung flacher Übergänge in der Standardtechnologie	11
2.2.1	Flache Übergänge in der Grundstruktur MOSFET	11
2.2.2	Skalierungseffekte und Ableitung von Anforderungen	12
2.2.3	Vorgaben für flache Übergänge in der ITRS-Roadmap	17
2.3	Verbreitete Herstellungsverfahren von flachen Übergängen und deren Beschränkungen	20
2.3.1	Ionenimplantation und thermisches Ausheilen	20
2.3.2	Alternative Dotierverfahren	22
2.4	Wahl des Dotierverfahrens	25
2.5	Zusammenfassung	26
3.	Grundlagen zum Dotierprozess mit Phosphor aus einer Oberflächenquelle	27
3.1	Phosphor für die Dotierung von Silicium	27
3.1.1	Eigenschaften von Phosphor	27
3.1.2	Eigenschaften von Phosphin	28
3.1.3	Anlagerung von Phosphin an der reinen Siliciumoberfläche	29
3.1.4	Festkörperlöslichkeit von Phosphor in Silicium	30
3.2	Diffusion von Phosphor in Silicium	32
3.2.1	Beschreibung der Diffusion nach Fick	32
3.2.2	Vergleich der Profilform – Abschätzung der benötigten Dosis	35
3.2.3	Diffusionskoeffizienten von Phosphor in Silicium	36
4.	Experimentelle Vorgehensweise	40
4.1	RTP-Modul	40
4.2	Eingesetzte Materialien	45
4.2.1	Verwendete Gasgemische	45
4.2.2	Substratauswahl und -vorbehandlung	46
4.3	Prozessführung - Auswahl der Experimente	47
4.3.1	Untersuchungen zur Reinigung und Vorbehandlung der Wafer	47
4.3.2	Temperaturkalibrierung der Anlage	48
4.3.3	Dotierexperimente	52
4.4	Charakterisierung der dotierten Siliciumwafer	53
4.4.1	Untersuchungen zur Belegung der Oberfläche	53
4.4.2	Ermittlung von Siliciumeigenschaften nach der Dotierung	62

5.	Ergebnisse und Diskussion.....	73
5.1	Vorbehandlung der Siliciumwafer	73
5.1.1	Einschätzung der ellipsometrischen Messungen	73
5.1.2	Vergleich der Reinigungsschritte	75
5.1.3	Stabilität der Oberfläche und Oxidwachstum an Luft	77
5.1.4	Analyse der Spektren.....	78
5.1.5	Schlussfolgerungen	82
5.2	Temperaturkalibrierungen.....	83
5.2.1	Zusammenhang von Temperatur und Lampenleistung.....	83
5.2.2	Temperaturkalibrierungen über mit Thermoelement bestücktem Wafer.....	87
5.2.3	Untersuchungen zur Homogenität der Temperaturverteilung der Anlage.....	90
5.2.4	Schlussfolgerungen	98
5.3	Einstufige Gasphasendotierung.....	99
5.3.1	Dotiergas hoher Konzentration.....	99
5.3.2	Dotiergas geringer Konzentration	105
5.3.3	Schlussfolgerungen	107
5.4	Zweistufiger Dotierprozess.....	107
5.4.1	Verfahren, Ablauf und angewendete Effekte	107
5.4.2	Prozessanalyse: Dotiererfolg, Homogenität und Reproduzierbarkeit.....	109
5.4.3	Flächenwiderstand in Abhängigkeit von Prozessparametern	118
5.4.4	Flächenwiderstand und Oxidation	121
5.4.5	Dotierstoffverteilung im Silicium.....	128
5.4.6	Untersuchungen der Belegung.....	142
5.4.7	Oberflächeneigenschaften der dotierten Wafer	144
6.	Zusammenfassung und Ausblick.....	152
7.	Danksagungen	155
8.	Liste der verwendeten Formelzeichen.....	156
9.	Liste der verwendeten Abkürzungen.....	159
10.	Abbildungsverzeichnis	161
11.	Tabellenverzeichnis.....	167
12.	Literatur.....	169

1. Einleitung

Bedeutende Fortschritte in Technik und Gesellschaft sind verbunden mit der rasanten Entwicklung der Informationstechnologie. Die sozialen und kulturellen Folgen erstrecken sich auf nahezu alle Bereiche des täglichen Lebens: auf Kommunikation, Wirtschaft, Industrie, Logistik, Bildung und Gesundheitswesen, um nur einige zu nennen. Das Internet revolutioniert den Informations- und Wissensaustausch. Der Umgang mit elektronischen Geräten ist selbst im Alltag zu einer Selbstverständlichkeit geworden, offensichtlich für viele durch die Nutzung von Computern, mobilen Kommunikationsgeräten oder Heim-elektronik, jedoch oft auch im Verborgenen von Kraftfahrzeugen oder Steuerungen und Regelungen intelligenter Geräte des täglichen Gebrauchs. Eine wichtige Grundlage für die Funktion dieser Geräte ist nach wie vor in herausragendem Maße das Material Silicium mit seinen besonderen physikalischen Eigenschaften sowie eine komplexe und innovative Technologie für dessen Bearbeitung. Diese ist letztendlich das Ergebnis einer mit ihrer fortschreitenden Verbesserung wachsenden Zahl von Forschungs- und Entwicklungsarbeiten, welche durch ihre Beiträge diesen Innovationsprozess ermöglichten und auch weiterhin gestalten werden.

1.1 Motivation und Zielsetzung der Untersuchungen zur Siliciumdotierung

1.1.1 Einordnung in die Siliciumtechnologie

Eine bedeutende Industrie hat sich in wenigen Jahrzehnten auf der Grundlage des Halbleitermaterials Silicium entwickelt mit einem stetigen Bedarf an Innovationen, die ein außerordentliches Wachstum schufen und weiterhin ermöglichen. Ein Meilenstein der Entwicklung war die Erfindung der integrierten Fertigung von Halbleiterbauelementen durch Jack Kilby 1958, der für seine Beiträge zur Erfindung von integrierten Schaltkreisen im Jahre 2000 den Nobelpreis erhielt. Ausdruck für die enorme Entwicklung seither ist das Mooresche Gesetz. Aus der Analyse des Wachstums von 1959 bis 1965 konnte Moore 1965 [*Moore '65*] eine jährliche Verdopplung der Anzahl der Komponenten pro integrierter Schaltung für die künftige Entwicklung ableiten. Dieses empirische Gesetz stellt bis heute (eine Anpassung erfolgte im Jahre 1975 [*Moore '75*] auf eine Verdopplung alle 2 Jahre) die Wegleitung einer wachsenden Industrie dar. Die weitere Erfüllung dieses Gesetzes erfolgt durch Planungen und Zielstellungen der Halbleiterindustrie selbst, festgelegt in der International Technology Roadmap for Semiconductors [*ITRS '06*]. Daraus leiten sich die Notwendigkeit und die Herausforderungen zukünftiger Entwicklungsrichtungen ab, mit einem stetigen Bedarf an Innovationen.

Eine zentrale Technologie für die Herstellung von integrierten Schaltungen (IC - integrated circuit) ist die CMOS-Technologie (CMOS - complementary metal oxide semiconductor), deren Name sich aus den Grundbauelementen, den MOS-Feldeffekttransistoren herleitet, die in den beiden komplementären Formen PMOS und NMOS, mit p- bzw. n-leitenden Kanalgebieten, zum Einsatz kommen.

Die Verwendung dieser Bauelemente und ihre fortschreitende Miniaturisierung, das heißt die Herabskalierung ihrer Dimension, bedingen zu einem wesentlichen Anteil die Anforderungen an die Technologie und deren künftige Entwicklung. Die Skalierung der CMOS-Bauelemente ist für eine weitere Erhöhung der Leistungsfähigkeit der integrierten Schaltungen erforderlich aber auch in besonderem Maße für eine Erhöhung der Produktivität und des wirtschaftlichen Ertrages.

Beim gegenwärtigen Stand der Miniaturisierung mit dem damit verbundenen großen Umfang an integrierbaren Komponenten auf einem elektronischen Bauteil wird die zur dessen Fertigung erforderliche Technologie als VLSI (very large scale integration) oder sogar ULSI (ultra large scale integration) bezeichnet.

Die Leistungssteigerung zeigt sich in erster Linie in einer Erhöhung der Schaltgeschwindigkeiten, was zu höheren Arbeitsgeschwindigkeiten der ICs führt, und in einer erhöhten Funktionalität durch die Integration einer größeren Anzahl von Bauelementen pro IC. Dadurch sinken auch die Kosten je Funktion bzw. für die Geschwindigkeit der hochintegrierten Bauelemente als Maß für die gestiegene Wirtschaftlichkeit. Dabei sollte nicht vergessen werden, dass die eigentlichen Triebkräfte der stürmischen Entwicklung gerade diese ökonomischen Gründe sind.

Als Maß für die Skalierungen kann ein Skalierungsfaktor K angegeben werden, mit dem sich interessante und weit reichende Konsequenzen der Miniaturisierung ableiten lassen (siehe z.B. in [Iwai '99]). Wenn die Taktfrequenz um den Faktor K steigen soll und die Anzahl der Transistoren pro IC-Fläche um den Faktor K^2 , müssen beispielsweise die Gatelänge L_G aber auch die Gateoxidicke t_{ox} und die pn-Übergangstiefe z_j um $1/K$ verringert werden. Es lässt sich also ableiten, dass mit der lateralen Skalierung der Bauelemente auch eine vertikale Skalierung erforderlich ist. All diese Skalierungen stellen jedoch stetig wachsende Anforderungen an die Technologie der Fertigung dar. Das Aufzeigen dieser Anforderungen und der damit verbundenen Schwierigkeiten und Probleme der Herstellung erfolgt in der ITRS-Roadmap, ein Entwicklungstempo vorgebend, das an das Mooresche Gesetz angelehnt ist. Die Roadmap stellt damit gewissermaßen den strategischen Plan der Halbleiterindustrie dar; ist darüber hinaus jedoch auch ein Mittel, das das Mooresche Gesetz zu einer sich selbst erfüllenden Prophezeiung macht.

In dem Maße jedoch, wie die Fertigungsverfahren immer komplexer, aufwendiger und damit teurer werden, steigen auch die Anforderungen an die Forschungs- und Entwicklungsarbeit. Auch hier wächst der Bedarf an Mitteln und Ressourcen, die zum Teil nur noch von den großen Herstellern oder Forschungsverbänden erbracht werden können. Wichtige Entwicklungsschwerpunkte und -richtungen können hier verfolgt werden. Aufgrund der großen Zahl einzelner Verfahrensschritte im Verband der Gesamttechnologie existieren darüber hinaus jedoch viele Möglichkeiten, auch abweichend von den vorherrschenden Technologien und Entwicklungsschwerpunkten, spezielle Felder zu bearbeiten und mit neuen Impulsen zu beleben.

1.1.2 Arbeitsschwerpunkte

Wie bereits gezeigt, ist ein Aspekt der Miniaturisierung auch die Verringerung der Tiefe der pn-Übergänge. Die vorliegende Arbeit befasst sich mit einer Methode zur Schaffung von flachen Übergängen in Silicium, insbesondere mit der n-Dotierung durch Phosphor.

Sie wurde untersucht

- als alternatives Verfahren zur Phosphordotierung zur Vermeidung und Verringerung von Schwierigkeiten, die bei konventionellen Verfahren auftreten,
- als einfaches Verfahren, das ein günstigere Verfahrensweise gestattet und
- als Möglichkeit, physikalische Vorgänge wie die Diffusion von Phosphor in Silicium oder auch Oberflächeneigenschaften von Silicium zu studieren.

Konkret bestand die Aufgabenstellung darin, ein Dotierverfahren zu finden und zu untersuchen, das als Alternative zu den in der Industrie für die Dotierung flacher Übergänge verwendeten Implantationsverfahren geeignet ist. Für solche Verfahren müssen in der Regel sehr kostspielige Anlagen eingesetzt werden. Grundsätzlich sind jedoch nicht für alle Dotierprozesse diese Implantationen erforderlich, die darüber hinaus auch einige prinzipielle Nachteile aufweisen. Es wurde nun angestrebt, unter Einsatz von Phosphin, einer gasförmigen Phosphorquelle, in einer konventionellen Kurzzeitprozessanlage flache pn-Übergänge in Silicium zu erzeugen. Die Möglichkeiten und Beschränkungen diese Anlage waren dabei kritisch einzuschätzen. Dabei lag der Schwerpunkt der Arbeit nicht bei der Fertigung von Bauelementen, sondern in der Auswahl und Untersuchung des Dotierprozesses selbst und auf der Charakterisierung des dotierten Siliciums.

Die Anwendung dieser Prozesse liegt an erster Stelle in der Dotierung von einkristallinen Siliciumschichtbereichen zur Erzeugung flachster pn-Übergänge im Bereich von ca. 10 nm bis 70 nm Tiefe. Der verfolgte Ansatz eines Dotierungsprozesses mit einer Oberflächenquelle aus einer definierten Belegung mit Phosphor soll somit eine alternative, wenig aufwendige und strukturdefektfreie Methode insbesondere zur Herstellung von Source-/Drain-Extension-Bereichen von Silicium-Halbleitertransistoren darstellen.

1.2 Gliederung der Arbeit

Im folgenden zweiten Kapitel wird die Dotierung von Silicium zur Erzeugung flacher pn-Übergänge als wichtiger Aspekt in der gegenwärtigen Technologie der Fertigung von Silicium-Bauelementen vorgestellt. An dieser Stelle wird die Bedeutung der im Rahmen der vorliegenden Arbeit durchgeführten Untersuchungen abgeleitet und im Zusammenhang zu gegenwärtigen Verfahren und Entwicklungen aufgezeigt. Dazu werden verbreitete Verfahren zum Dotieren von Silicium für flache pn-Übergänge analysiert und die Motivation für die Auswahl des untersuchten Dotierverfahrens für eine Phosphordotierung dargelegt.

Im dritten Kapitel werden die Grundlagen für eine Phosphordotierung erläutert, die für die untersuchten Prozesse relevant sind. Wichtige Eigenschaften von Phosphor und Phosphin werden vorgestellt. Dabei werden besonders deren Wechselwirkungen bei Kontakt mit dem Silicium-Halbleitermaterial berücksichtigt. Die Diffusion von Phosphor in Silicium, die für die Beschreibung der erwarteten Transportprozesse bei der Dotiertemperatur erforderlich ist, wird kurz erläutert, und aus den Gesetzmäßigkeiten werden Einflüsse von grundlegenden Prozessparametern auf die Dotierergebnisse abgeschätzt.

Im vierten Kapitel werden die für die experimentellen Arbeiten verwendeten Anlagen und Methoden vorgestellt und die Motivation und Durchführung von Experimenten erläutert. Außerdem werden hier die für die Charakterisierung der Wafer angewendeten Untersuchungsverfahren, Ellipsometrie, Sekundärionen-Massenspektrometrie und Vierspitzen-

messungen, die für die Auswertung und Bewertung der Ergebnisse von grundlegender Bedeutung sind, eingehend diskutiert.

Den Hauptteil der Arbeit stellt das fünfte Kapitel dar. Hier erfolgen die Vorstellung der Ergebnisse der experimentellen Untersuchungen und deren eingehende Diskussion.

Zunächst werden Untersuchungen zur Vorbehandlung der Wafer ausgewertet. Der Einfluss von Reinigungsverfahren auf die Oberflächeneigenschaften der Wafer wird analysiert. Ein Schwerpunkt liegt dabei in der Bewertung der Möglichkeiten einer einfachen und zerstörungsfreien Charakterisierung mittels ellipsometrischer Messungen. Diese Untersuchungen erfolgen zudem im Hinblick auf einen Vergleich mit Ergebnissen von Messungen an den prozessierten Wafern.

Eine besondere Bedeutung besteht in der Auswertung der Möglichkeiten und Beschränkungen der Prozessanlage hinsichtlich einer Erfassung und Kontrolle der Wafertemperatur während der Kurzzeitprozesse. Dies ist wesentlich für die Einschätzung der Dotierprozesse anhand der Messergebnisse. Von Bedeutung für die Nutzung der untersuchten Verfahren in Fertigungsprozessen sind dabei die Homogenität der Dotierergebnisse bezüglich der gesamten Waferfläche und die Stabilität der Prozesse für reproduzierbare Resultate.

Die Ergebnisse aus verschiedenen Dotierprozessen unter Einsatz von Phosphin in der Gasphase als Dotierstoffquelle werden vorgestellt. Die Beschränkungen des konventionellen Gasphasendotierprozesses werden aufgezeigt und ein zweistufiger Prozess vorgestellt, der eine Phosphordiffusion unter minimalem Gaseinsatz ermöglicht. Der Einfluss von wichtigen Prozessparametern auf die Dotierergebnisse wird gezeigt und aus den Ergebnissen auf die Möglichkeiten und Grenzen des Prozesses unter Berücksichtigung der Beschränkungen der verwendeten Anlage hingewiesen. Grundlegende physikalische Vorgänge im Verlauf und als Ergebnis der Dotierprozesse werden diskutiert und die Methoden der Charakterisierung der flachen pn-Übergänge bewertet.

Im abschließenden Kapitel erfolgen eine Zusammenfassung der wichtigsten Ergebnisse im Hinblick auf deren Bedeutung für die Prozesstechnologie, die Bewertung der Möglichkeiten einer Anwendung der untersuchten Prozesse sowie Vorschläge für weitere Arbeiten.

2. Dotierung von Silicium zur Erzeugung flacher pn-Übergänge

Die physikalischen Eigenschaften des Festkörpermateriale Silicium aber auch dessen praktisch unbegrenzte Verfügbarkeit bilden die Grundlage für seine Verwendung in elektronischen Festkörperbauelementen. Der Festkörper wird im idealen (einkristallinen) Fall aus Siliciumatomen gebildet, die in regelmäßiger Anordnung ein kubisch-flächenzentriertes Gitter ausbilden, wobei ein Gitterplatz von zwei benachbarten Atomen besetzt ist. Diese Anordnung entsteht durch die Bindung der 4-wertigen Atome zu jeweils vier Nachbarn, die dieses so umgeben, dass sie die Eckpunkte eines regelmäßigen Tetraeders bilden. Infolge der Überlagerungen der Wellenfunktionen der Einzelatome im Festkörper entsteht die besondere elektronische Struktur. Eine Modellierung dieser Struktur im einfachen Bändermodell der Elektronen ist eine geeignete Methode zur Beschreibung der wichtigsten elektrischen Eigenschaften. Die Ausbildung einer Bandlücke, die bei Silicium bei 300 K 1,12 eV beträgt [Madelung '91], bedingt die charakteristische mit der Temperatur steigende Leitfähigkeit.

Für die technologische Nutzung in Halbleiterbauelementen ist mindestens ebenso von Bedeutung, dass diese Eigenschaften durch Beimengung von geeigneten Fremdatomen in die Kristallstruktur gezielt beeinflusst werden können. Diese Dotierung erfolgt in Silicium durch Einbau von Elementen der 3. Hauptgruppe, das heißt durch Akzeptoren wie Bor, zur Bildung von p-leitenden Bereichen: Elektronenaufnahme durch die Boratome führt zu einem Elektronenmangel im Valenzband, wodurch eine Stromleitung über unbesetzte Zustände, so genannte Löcher, möglich wird. Entsprechend erfolgt die Dotierung von n-leitenden Bereichen durch Einbau von Elementen der 5. Hauptgruppe, den Donatoren wie Phosphor. Hier kommt es durch die zusätzlichen Elektronen, die nicht für die direkten Bindungen im Festkörper erforderlich sind, zur Besetzung neuer Zustände im Leitungsband, wodurch eine Stromleitung über die Elektronen erfolgen kann.

Räumlich benachbarte p- und n-leitende Bereiche im Festkörper, das heißt benachbarte Gebiete unterschiedlicher Dotierung, bilden eine besondere elektronische Struktur, den pn-Übergang, eine Grundstruktur der elektronischen Halbleiterbauelemente. Die unterschiedlichen Dotierungen führen zur Ausbildung von Raumladungszonen, die durch äußeres Anlegen von Spannungen oder durch Ladungsträgerinjektionen beeinflusst werden können und die Funktion der Halbleiterbauelemente wesentlich bestimmen.

2.1 Die Dotierung im Laufe der Entwicklung der Halbleitertechnologie

Die Entwicklung und Verfeinerung geeigneter Dotiermethoden ist eng verknüpft mit den Fortschritten in der Technologie der Herstellung von Halbleiterbauelementen. Obwohl viele physikalische Grundlagen der Halbleiterelektronik bereits seit den 30er Jahren des vorigen Jahrhunderts bekannt waren und Vorschläge für die Nutzung in Halbleiterbauelementen abgeleitet wurden, gelang die Erfindung des Transistors erst im Jahre 1947 in den Bell Laboratorien durch Schaffung eines Punkt-Kontakt-Überganges im polykristallinen Germanium. Dies kann als Meilenstein in der Geschichte der Technik aufgefasst werden; die weiteren Entwicklungen führten in den 50er Jahren zum Bipolartransistor, wofür J. Bardeen, W. Brattain und W. Shockley bereits 1956 den Nobelpreis erhielten.

Diese ersten Bauelemente waren noch aus Germanium, jedoch schon kurze Zeit später wurden Bauelemente aus Silicium vorgestellt. Bis zur Mitte der 50er Jahre wurden die

Übergänge noch weiterhin durch Einlegieren von Dotiermaterial aus Oberflächenschmelzen oder durch geeignete Dotierung der Schmelzen bei der Fertigung des Halbleitermaterials selbst hergestellt. Daraus konnten Einzelbauelemente hergestellt werden, eine Integration war mit diesen Fertigungsverfahren jedoch nicht möglich. Einen großen Fortschritt bedeutete dann die Einführung der Gasphasendiffusion. Die Dotierung erfolgte nun durch Aufheizen einer Scheibe aus Halbleitermaterial (Wafer) in einer Atmosphäre, die die einzubringenden Fremdatome in gasförmigen Verbindungen, so als Hydride wie Diboran (B_2H_6), Phosphin (PH_3) oder Arsin (AsH_3), enthielt. Nach Adsorption der Dotierstoffe an der Oberfläche erfolgte die Dotierung durch Diffusion im Festkörper. Dieser Schritt konnte mit anderem Dotierstoff für die Schaffung der gewünschten Übergänge parallel zur Oberfläche wiederholt werden. Nach Kontaktieren und Strukturieren zur so genannten Mesa-Bauweise, wobei durch Ätzen der Oberfläche erhabene und somit räumlich getrennte Strukturen einzelner aktiver Bauelemente geschaffen wurden, konnten so viele Bauelemente aus einem Wafer gewonnen werden. Nachteilig war jedoch weiterhin, dass sich die Übergänge bis zur Oberfläche oder zu den Kanten der einzelnen Bauelemente erstreckten.

In den späten 50er Jahren wurde dann der planare Fertigungsprozess entwickelt. Zu der Dotierung durch die Gasphasendiffusion zur Schaffung p- und n-leitender Bereiche kam nun die Möglichkeit der Strukturierung dieser Bereiche durch Maskierung mittels einer Oxidschicht. Silicium war nun vollends das Halbleitermaterial der Wahl für die Massenfertigung, da es die herausragende Eigenschaft besitzt, durch Oxidation stabile und elektrisch isolierende Oxidschichten zu bilden. Für den Dotierprozess und die Maskierung ist besonders von Bedeutung, dass Siliciumoxidschichten eine Diffusionsbarriere für die zur Dotierung verwendeten Dotierstoffe darstellen, das heißt, die Diffusionsraten der Dotierstoffe sind im Oxid bedeutend geringer als im Silicium selbst. Durch die Maskierung der aktiven Bereiche erfolgte nun auch gleichzeitig eine Passivierung an der Oberfläche. Dies bedeutet, dass nun die aktiven Übergänge unter einer Oxidschicht enden und sich nicht mehr bis zur Oberfläche erstrecken, so dass die elektrischen Eigenschaften der Übergänge nicht mehr durch Oberflächeneffekte gestört werden können.

Damit waren die Grundlagen gegeben, die zur Erfindung der integrierten Bauelemente 1959 durch J. Kilby führten und zu den photolithographischen Verfahren zur Herstellung integrierter MOS-Schaltkreise durch R. Noyce. Durch den Einsatz von Masken und die Strukturierung konnten nun mehrere Bauelemente und deren elektrische Verbindungen zu kompletten elektrischen Schaltungen, den ICs, auf einem kleinen Bereich des Wafers, dem Chip, zusammengefasst werden. Mit der darauf folgenden Einführung der CMOS-Technologie haben sich die grundlegenden Fertigungsmethoden zur Herstellung integrierter Bauelemente seit den 60er Jahren des vorigen Jahrhunderts nicht mehr wesentlich geändert. Die weitere Entwicklung ging evolutionär voran, aufgrund des Potentials dieser Technologie jedoch mit außerordentlichem Tempo und eindrucksvoll beschrieben durch das Mooresche Gesetz.

So ist auch mit der planaren Ausführung der aktiven Bereiche die eigentliche Schaffung der Übergänge durch Diffusionsprozesse im Wesentlichen geblieben, jedoch sind die Methoden zum kontrollierten Einbringen der Dotierstoffe in den erforderlichen Mengen in das Silicium erweitert worden. Die bis in die 70er Jahre allein gebräuchlichen Mittel waren die erwähnte Gasphasendotierung oder die Dotierung aus Glasschichten, die als Dotierstoffquelle auf der Siliciumoberfläche abgeschieden oder durch Oxidation in dotierstoffhaltiger Atmosphäre erzeugt wurden. Damit war jedoch eine von der Diffusion unabhängige Kontrolle der einzubringenden Dotierstoffmenge schwierig, da sie bei solchen Prozessen in der Praxis für die Erzielung reproduzierbarer Ergebnisse nur über die Festkörperlöslichkeit des Dotierstoffs im Silicium bei der Prozesstemperatur eingestellt werden konnte. Mitte der 70er Jahre wurde daher die Ionenimplantation in die Bauelementfertigung eingeführt und setzte sich als Standard für die Vorbelegung durch; sie erlaubte eine bessere Dosiskontrolle und -variation.

Die im Folgenden beschriebenen Anforderungen an flachste Profile machen eine weitere kontinuierliche Verbesserung der Dotierprozesse notwendig. Die Lage der dotierten Bereiche und die Dotierstoffkonzentration müssen immer genauer definiert werden können, wofür die Beherrschung und das Verständnis der Diffusionsprozesse, aber auch von Oberflächen- und Grenzflächenvorgängen, erforderlich sind, um die angestrebten elektrischen Eigenschaften zu erhalten.

2.2 Bedeutung flacher Übergänge in der Standardtechnologie

2.2.1 Flache Übergänge in der Grundstruktur MOSFET

Seit der Einführung der MOS-Technologie für die Herstellung von integrierten Bauelementen stellt der MOSFET (metal oxide semiconductor field effect transistor) die wichtigste Grundstruktur im Fertigungsprozess dar. Aus dessen Aufbau, den Design- und Leistungsanforderungen leiten sich die wesentlichen Anforderungen an die Fertigungstechnologie und somit auch an die Dotiertechnologie ab.

Der planare MOSFET¹ (siehe Abbildung 2.1) besteht aus den hochdotierten Bereichen Source (Quelle) und Drain (Senke), eingebettet in das Siliciumsubstrat und voneinander getrennt durch das Kanalgebiet, einen schmalen Bereich gegensätzlicher Dotierung. Flachere Bereiche hoher Dotierung, die Source-/Drain-Erweiterungen (SDE – source/drain extensions), verbinden Source und Drain mit dem Kanal. Über dem Kanal befindet sich eine Oxidschicht, die die Gate-Elektrode (Steuerelektrode) vom Kanalbereich elektrisch isoliert. Die Source- und Draingebiete sind durch Silicidkontakte mit der Schaltung verbunden. Die Gate-Elektrode besteht gewöhnlich noch aus dotiertem Polysilicium, das auch über ein Silicid kontaktiert ist. Die Gate-Elektrode ist von Spacern umgeben, Oxidbereichen, die in der Technologie der selbstjustierenden Masken benutzt werden. Das einzelne Bauelement ist zu benachbarten Elementen durch dicke Oxide oder Grabenstrukturen (Trenches), die mit Oxiden gefüllt sind, abgegrenzt.

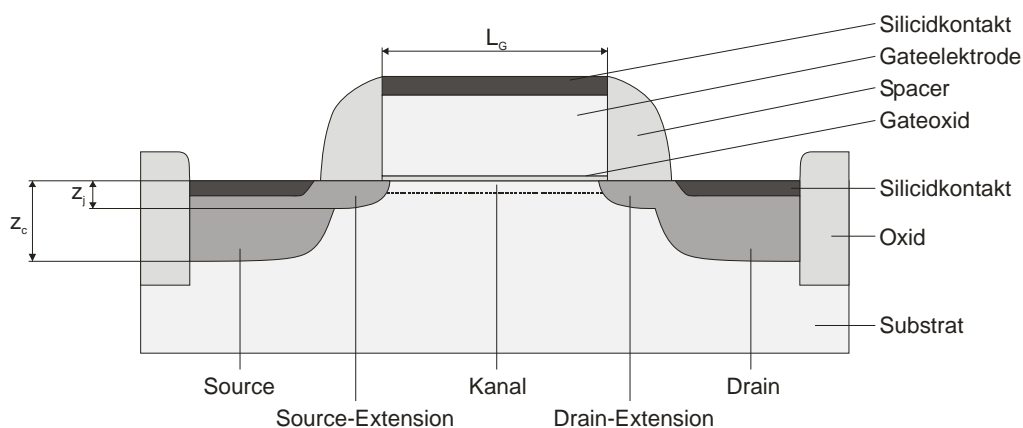


Abbildung 2.1: Schematische Darstellung eines n-Kanal MOSFET.

¹ An dieser Stelle soll nur die konventionelle Bauform beschrieben werden, da dies für die weiteren Ausführungen der Anforderungen an die Dotierprozesse genügt. Weitere Bauformen, die zum Teil noch in der Entwicklung sind, wie SOI, FinFET oder vertikale Strukturen sind in der Literatur beschrieben, siehe z.B. [Hutchby '02; Wong '02; Huff '04; King '05].

Der Feldeffekttransistor ermöglicht, mittels einer Gatespannung U_G , einen bei Anlegen einer Spannung U_{SD} zwischen Source und Drain erfolgenden Stromfluss, den Drainstrom I_D , zu steuern. Liegt keine Spannung am Gate an oder ist sie geringer als die Einsatzspannung U_{th} (threshold voltage), besteht keine leitfähige Verbindung zwischen Source und Drain, und eine Potentialbarriere eines in Sperrrichtung geschalteten pn-Übergangs verhindert im idealen Fall für den hier betrachteten gewöhnlichen Anreicherungstyp-FET einen Stromfluss I_D selbst bei Anlegen von U_{SD} . Im realen Bauelement ist jedoch selbst im Aus-Zustand ein geringer, unerwünschter Stromfluss nicht vermeidbar, und ein Leckstrom (sub-threshold leakage current) fließt so zwischen Source und Drain. Durch Anlegen einer Gatespannung, die die Einsatzspannung übersteigt, kann der Transistor eingeschaltet werden, es kommt zur Ausbildung eines leitfähigen Kanals zwischen Source und Drain durch Induzierung einer Inversionschicht unterhalb des Gateoxids. Die Leitfähigkeit des Kanals ist der angelegten Gatespannung direkt proportional. Bei Erhöhung der Source-Drain-Spannung und einer Gatespannung oberhalb von U_{th} tritt eine Sättigung des Drainstroms auf.

Aufgrund der Möglichkeit einer planaren Bauweise mit sehr geringen Leckströmen im Aus-Zustand und wegen des geringen Stromverbrauchs des isolierten Gates für die Schaltvorgänge, wurde der MOSFET zum Grundbaustein der integrierten Schaltungen.

In diesem Bauelement sind nun ohne Anlegen von äußeren Spannungen pn-Übergänge sowohl zwischen den Source- als auch den Drainbereichen und dem jeweiligen benachbarten Substratgebiet ausgebildet. Im Bereich der Source-/Drain-Extensions, die sich nur bis in sehr geringe Tiefen erstrecken dürfen, wie im folgenden näher erläutert werden wird, liegen nun diese Übergänge besonders nahe an der (ursprünglich planaren) Oberfläche und werden daher auch als flache oder auch ultra-flache Übergänge (USJ - ultra shallow junction) bezeichnet. Ihrer vertikalen Dimension bezüglich einer horizontalen Ober- oder Grenzfläche kommt bei der Charakterisierung von Eigenschaften bei der planaren Bauweise eine besondere Bedeutung zu, so ist hier die Tiefe ein wichtiges Kriterium für die Übergänge. Diese Übergangstiefe z_j bezieht sich nun im eigentlichen Sinne auf die Stelle, an der die Dotierstoffkonzentration des eingebrachten Dotierstoffes auf den Wert der entgegengesetzten Substratdotierung fällt, dies wird auch als die Stelle des metallurgischen Übergangs bezeichnet. Hier sei jedoch bereits darauf verwiesen, dass oft, insbesondere in Untersuchungen zur Ausbreitung von Dotierstoff im Substrat, das dann nicht einmal gegensätzliche Dotierung aufweisen muss, eine „Übergangs“-Tiefe vielmehr *definiert* wird, als die Stelle, bei der eine bestimmte Dotierstoffkonzentration unterschritten wird. Diese Konzentration wird so gewählt, dass damit den Dotierverhältnissen in den realen Bauelementen im Halo- und Wannbereich Rechnung getragen wird. In dieser Arbeit soll, wenn nicht anders bezeichnet, die Übergangstiefe auf eine Konzentration von $1 \cdot 10^{18} \text{ cm}^{-3}$ bezogen werden.

2.2.2 Skalierungseffekte und Ableitung von Anforderungen

Die angestrebte laterale Skalierung des MOSFET zur Erhöhung der Packungsdichte kann nun am besten so erfolgen, dass die Gatelänge L_G reduziert wird, da dies auch zu einem geringeren Widerstand im Kanalgebiet und durch Verringerung der Kapazitäten zu kürzeren Schaltzeiten führt. Wenn jedoch die Gatelänge mit $1/K$ skaliert wird (Skalierfaktor $K > 1$), sollten für ein entsprechendes Funktionieren des Bauelements die anderen Dimensionen idealerweise mit skaliert werden, um bei skaliertem Betriebsspannung gleiche Feldverhältnisse zu erhalten und das elektrische Feld auf die dünne Kanalregion unterhalb des Gates zu beschränken. Somit ergibt sich bereits aus den Forderungen nach Miniaturisierung und Leistungssteigerung auch die Anforderung an eine *vertikale* Skalierung und damit auch an die Tiefe der pn-Übergänge.

2.2.2.1 Kurzkanaleffekte

Eine weitere Herausforderung besteht dadurch, dass im Zusammenhang mit der Reduzierung der Gatelänge besondere Effekte auftreten, die die elektrischen Eigenschaften des MOSFETs stören und verschlechtern. Diese Kurzkanaleffekte (SCE - short channel effect) sollen im Folgenden kurz erläutert werden. Es zeigt sich, dass die Reduzierung und genaue Kontrolle der Übergänge insbesondere im Bereich der SDE ein wichtiges Mittel ist, die Kurzkanaleffekte zu minimieren, da sie direkt oder indirekt von den Dimensionen der Source-/Drain-Übergänge abhängen. Prinzipiell könnten die Kurzkanaleigenschaften auch durch größere Source-/Drain-Übergangstiefen verbessert werden. Dies erforderte jedoch auch eine höhere Substratdotierung, was einerseits erhöhte Substratkapazitäten mit dadurch höheren Schaltzeiten und einer vom Substratbias abhängigen Einsatzspannung zur Folge hätte und außerdem zu höheren Leckströmen führte (siehe z.B. [Jones '98], auch für die folgenden Ausführungen).

Verschiebung der Einsatzspannung

Für ideale MOSFETs mit langen Kanälen im Vergleich zu den Übergangstiefen von Source und Drain ist die Einsatzspannung U_{th} nur abhängig von der Gatekapazität, der Kanal-dotierung und der Austrittsarbeit des Materials der Gate-Elektrode. In den realen Bauelementen wird jedoch beobachtet, dass bei Reduzierung der Gatelänge die für die Ausbildung eines leitfähigen Kanals notwendige Einsatzspannung geringer wird und gegen Null geht. Das heißt, dass damit eine Kontrolle des Einschaltens des MOSFETs durch das Gate verringert wird oder ganz verloren geht. Dieser Kurzkanaleffekt wird auch threshold voltage roll-off genannt. Der Grund dafür ist, dass sich die Raumladungszonen der Source-/Drain-Übergänge in den Kanalbereich unter das Gate erstrecken und diese Verarmungszone nicht mehr als einfach abrupt rechteckförmig angenommen werden kann.

Barriereerniedrigung durch Drain-Einfluss

Wenn außerdem, bei gesperrtem pn-Übergang am Drain, die Drainspannung erhöht wird, vergrößert sich die Verarmungszone am Drainübergang weiter in den Kanal hinein. Damit vermindert sich die Potentialbarriere zwischen Source und Drain und somit auch die effektive Kanallänge, so dass, nun abhängig von der Drainspannung, die Einsatzspannung weiter verringert wird. Dieser Effekt fällt insbesondere bei kurzer Kanallänge stärker ins Gewicht und wird drain-induced barrier lowering (DIBL) genannt.

Durchgriff

Wenn sich die Verarmungszonen von Source und Drain unter dem Kanal soweit angenähert haben, dass sie sich treffen, spricht man vom Durchgriff (punch-through). In diesem Falle ist die Potentialbarriere für den Stromfluss gemindert, und außerhalb des Kanalgebietes, das heißt ohne die Kontrolle durch das Gate, können sich alternative Strompfade ausbilden. Bei gleich bleibender Dotierung des Kanalgebietes tritt dieser Effekt eher auf, je geringer der Abstand zwischen Source und Drain ist, das heißt je geringer die Kanallänge ist. Die Verringerung der Ausdehnung der Verarmungsgebiete von Source und Drain durch Erhöhung der Dotierung im Kanalgebiet oder im gesamten Substratbereich außerhalb von Source und Drain ist jedoch nicht möglich, ohne dass die Ladungsträgerbeweglichkeit im Kanal zu sehr verringert würde und verstärkt vom Substratbias abhängige Einsatzspannungen bzw. erhöhte Substratkapazitäten auftraten.

Lokale Dotierungen unterhalb des Kanals können eingesetzt werden, um den Durchgriff zu vermindern: tiefe Implantationen für abgestufte Wannendotierungen und Halo-Implantationen, um die Dotierung im Substrat lokal neben oder unter den Source-/Drain-Übergängen zu erhöhen. Eine Verbesserung wird hier auch durch die Trennung des aktiven Gebietes von Substrat durch isolierende Barrierschichten, wie mit der Bauweise Silicon-on-Insulator (SOI), erreicht.

Eine ultraflache Ausführung von SDE-Bereichen im Bulk-CMOS-Prozess wirkt sich günstig auf die Vermeidung des Durchgriffs aus, da somit die eigentlichen Source-/Draingebiete, die für eine gute Kontaktierung tiefer ausgeführt werden müssen (Kontakttiefe z_c), weiter, das heißt mehr als die Kanallänge, voneinander entfernt platziert werden können.

Verschlechterung des Abschaltverhaltens

Das Abschaltverhalten des MOSFETs wird durch den Verlauf der Kennlinie des Drainstroms I_D über U_G für Gatespannungen unterhalb der Einsatzspannung (sub-threshold) charakterisiert. In diesem Bereich ist zwar kein durchgehender, leitfähiger Kanal mehr ausgebildet, jedoch tritt noch ein Stromfluss auf, der überwiegend durch Diffusion bestimmt ist. Er ist abhängig von der Ladungsträgerkonzentration im Kanal, die von dem Oberflächenpotential über dem Kanal exponentiell abhängt. Es ergibt sich, dass für eine Verringerung von I_D um den Faktor 10, das heißt um eine Dekade, das Oberflächenpotential um 60 mV fallen muss. Bei idealem MOSFET mit langem Kanal ist dieses Potential vollständig vom Gate mit dessen Kapazität bestimmt, so dass eine Verringerung der Gatespannung um 60 mV auch eine Verminderung von I_D um eine Dekade bewirkt. Dies ist der so genannte sub-threshold-swing. Für Kurzkanalbaulemente sind jedoch die von der Gatespannung nicht direkt beeinflussbaren Beiträge der Kapazitäten von Source, Drain und Substrat auf die Ladungsträgerkonzentration im Kanal nicht mehr zu vernachlässigen, und für eine entsprechende Verringerung von I_D ist ein größerer Spannungsabfall als 60 mV am Gate erforderlich. Anders ausgedrückt bedeutet dies, dass bei gleicher Verminderung von U_G wie bei MOSFETs mit langen Kanälen noch ein höherer Strom I_D vorhanden ist. Damit erfolgt das Abschalten langsamer und der Leckstrom ist dabei höher. Durch geringere Source-/Drain-Kapazitäten kann dieser nachteilige Kurzkanaleffekt vermindert werden.

2.2.2.2 Auswirkung der Skalierung auf weitere MOSFET-Eigenschaften

Erhöhung des Serienwiderstands

Die Verminderung der Übergangstiefen von Source und Drain hat jedoch nicht nur positive Auswirkungen auf die gewünschten Bauelementeigenschaften. Der parasitäre Serienwiderstand zwischen Source und Drain im Ein-Zustand steigt mit flacheren Übergängen an. Dies hat eine Verringerung des Drainstroms zur Folge, was zu Leistungsverlust und Wärmeentwicklung führt.

Der Gesamtwiderstand des jeweiligen Source- und Draingebietes setzt sich zusammen aus einer Reihenschaltung

- vom Kontaktwiderstand des Silicidkontakts,
- vom Flächenwiderstand,
- vom Ausbreitungswiderstand zum Kanal und
- vom spannungsabhängigen Akkumulationswiderstand im dem Bereich, wo sich die SDE-Gebiete unter das Gate erstrecken.

Die Widerstände sind abhängig vom jeweiligen Dotierprofil, der Übergangstiefe und dem Aktivierungsgrad der Dotierstoffe.

Insbesondere der Flächenwiderstand R_S ist abhängig von der Dotierstoffverteilung. Er erhöht sich, wenn die Übergangstiefe z_j vermindert wird, es sei denn, die Ladungsträgerbeweglichkeiten μ_p , μ_n oder -konzentrationen p bzw. n erhöhten sich entsprechend:

$$R_S = \frac{1}{z_j \int_0^{z_j} q [\mu_p p(x) + \mu_n n(x)] dx} . \quad (1)$$

Jedoch tritt bei Erhöhung der Dotierkonzentrationen aufgrund von Ladungsträgerstreuungen an den Dotierstoffdefekten vielmehr eine Verringerung der Mobilität bis zur Sättigung auf einen geringeren Wert als für intrinsisches Material auf. Durch zusätzliche Defekte wie Dotierstoffpräzipitate oder Kristallfehler nach Implantationen wird die Mobilität noch weiter vermindert.

Der Akkumulationswiderstand wird besonders von den Halo-Dotierungen und dem lateralen Dotierverlauf der SDE beeinflusst. Abrupte Übergänge erweisen sich hier als günstig für einen geringen Widerstand, jedoch auch nachteilig in Bezug auf erhöhte Leckströme. Der Widerstand ist zudem spannungsabhängig, die speziellen Anforderungen an den lateralen Profilverlauf ergeben sich erst aus detaillierteren Simulationen von Bauelementen [Gossmann '03; Taylor '04].

Effekte durch energiereiche Ladungsträger

Im Bauelement können im Ein-Zustand, bei Anliegen einer hohen Source-Drain-Spannung im Sättigungsbereich von I_D sehr hohe elektrische Felder an den Drainkanten auftreten. Eine Folge davon sind energiereiche „heiße“ Ladungsträger (hot electrons), die zu Interface- und Oxidschäden führen können, was negativen Einfluss auf den Drainstrom, die Einsatzspannung und das Abschaltverhalten haben kann. Die maximale Feldstärke im Kanal ist umso größer, je geringer die Übergangstiefe im Drainbereich ist und je höher V_{DS} über der Spannung ist, bei der eine Sättigung von I_D auftritt.

Kurzkanal-Bauelemente mit flachen Übergängen und geringer Sättigungsspannung sind damit anfällig für diesen Effekt. In der Vergangenheit wurden aufgrund der Verwendung höherer Betriebsspannungen für die Bauelemente flache, gering dotierte Bereiche zwischen Drain und Kanal geschaffen (LDD - lowly doped drain), um den Spannungsabfall im Kanal zu verringern. Dies ist jedoch äußerst nachteilig in Bezug auf den Serienwiderstand, wie oben beschrieben. Wegen den heute üblichen geringeren Betriebsspannungen fällt dieser Effekt nicht mehr nachteilig ins Gewicht, so dass nun statt der gering dotierten LDD-Bereiche die hoch dotierten Source-/Drain-Extensions für eine Optimierung des Serienwiderstands einsetzbar sind.

Leckströme

Für kleinere Bauelemente mit geringeren allgemeinen Betriebsströmen ist die Vermeidung von parasitären Leckströmen von besonderer Bedeutung. Dies ist besonders wichtig für die Verminderung des Stromverbrauchs im Aus-Zustand und für Speicherbauelemente, die elektrische Ladungen speichern, die über die flachen Übergänge oder Leckströme abfließen könnten.

Da die Verarmungsbereiche bei flachen Übergängen näher an der Ober- oder Grenzfläche sind, können erhöhte Leckströme auch verstärkt durch Defekte an diesen Stellen auftreten, z.B. durch Oberflächendefekte aus der Prozessierung, durch Kontaminationen an der Oberfläche oder durch Stress aus Beschichtungen. Allgemein gilt, dass mit der Verringerung des Abstandes zu Grenzflächen im Zuge der Skalierung, diese auch eine zunehmend größere Rolle für das Verhalten der Bauelemente spielen.

Die Leckströme (Sperrströme) an einseitig hoch dotierten Übergängen setzen sich zusammen aus der Diffusion von Minoritätsladungsträgern durch die Verarmungszone und dem Strom aus Ladungsträgern infolge elektrischer Felder. Nach der Shockley-Read-Hall-Statistik ist die Nettogeneration als Quelle von Elektronen und Löchern in der Verarmungszone umso höher, je höher die angelegte Spannung ist. Mit verminderter Lebensdauer von Minoritätsladungsträgern und geringerer effektiver Lebensdauer der generierten Ladungsträger steigt der Sperrstrom, der für das stationäre Gleichgewicht der Ladungsträgerkonzentration erforderlich ist. Die Lebensdauer hängt nun entscheidend von der Anzahl der Rekombinationszentren der Minoritätsladungsträger und Generationszentren für den Generationsstrom ab, das heißt von der Anzahl der Defekte, wie tiefe Störstellen und Traps.

Für die flachen Übergänge ist es daher extrem wichtig, dass Defekte, wie sie z.B. bei der Implantation unvermeidlich sind, durch Hochtemperaturschritte (annealing) vollständig ausgeheilt werden und keine Generations- und Rekombinationszentren, sei es durch Implantation, Stress oder Kontaminationen, in den Verarmungszonen verbleiben. Als besondere Schwierigkeit ergibt sich jedoch hier, dass Hochtemperaturschritte äußerst ungünstig für die Schaffung flacher Dotierprofile sind, da die damit einhergehende Diffusion exponentiell von der Temperatur abhängig ist und darüber hinaus hohe Defektkonzentrationen die Diffusion außerordentlich beeinflussen können (siehe auch Abschnitt 3.2).

2.2.2.3 Anforderungen an den Fertigungsprozess

Aus den oben genannten physikalischen Effekten ergibt sich eine Reihe von Anforderungen an den Dotierprozess zur Erzeugung von flachen Übergängen. Im Vordergrund steht die Erzielung einer geringen Tiefe in vertikaler und einer geringen Ausbreitung in lateraler Richtung verbunden mit geringem Widerstand in der dotierten Schicht.

Wichtiges Kriterium neben der eigentlichen Übergangstiefe ist dabei auch die Form des Dotierstoffprofils, das heißt der Verlauf der Dotierstoffkonzentration im Siliciumsubstrat, da er einen wesentlich Einfluss auf den Widerstand besitzt. Bei idealem, abruptem, das heißt rechteckförmigem Dotierverlauf mit gleich bleibender Konzentration bis zum Übergang sind bei gegebener Dosis und Aktivierung von Dotierstoff der Akkumulations- und Flächenwiderstand minimal gegenüber allen anderen Verläufen bis zum Übergang bei gleicher Tiefe. Das bedeutet auch, dass bei gleicher Oberflächenkonzentration und Übergangstiefe derartig abrupte Profile gegenüber allmählich abfallenden von Vorteil sind, nicht nur in Bezug auf geringe Widerstände, sondern auch, weil sie ideale pn-Dioden mit geringerer Ausdehnung der Verarmungszone in den dotierten Bereich darstellen. Dies ist insbesondere bei den geringen Dimensionen der aktiven Bereiche zur Vermeidung von vollständiger Verarmung und Erhaltung möglichst idealer Diodeneigenschaften mit geringen Leckströmen erforderlich. Bei den realen Verläufen zeigt sich entsprechend, dass die gaußförmigen Verläufe den gewöhnlichen exponentiellen Verläufen bzw. denen entsprechend der komplementären Fehlerfunktion in dieser Hinsicht überlegen sind.

Der Fertigungsprozess für die Übergänge muss nun in die bestehende CMOS-Technologie mit einem möglichst geringen Aufwand integrierbar sein. Das heißt, es dürfen die übrigen Fertigungsschritte und -ergebnisse nicht gestört oder wesentlich beeinflusst werden, Defekte und Kontaminationen auf und im Substrat müssen vermieden werden, es sollten möglichst

bewährte Fertigungsanlagen genutzt werden können, und es muss ein hoher Durchsatz bei hoher Ausbeute und geringen Betriebskosten möglich sein.

Der Dotierprozess umfasst im Allgemeinen die Einbringung von Dotierstoff, die Aktivierung und die Diffusion. Dabei ist ein Kompromiss zwischen der Erzielung einer maximalen Aktivierung und einer geringstmöglichen Ausbreitung durch Diffusion zu erreichen. Die hohe Aktivierung, das heißt ein hoher Anteil von Dopanten auf Siliciumgitterplätzen, die Ladungsträger im Leitungs- bzw. Valenzband des Festkörpers zur Verfügung stellen, ist für geringe Serien- und Kontaktwiderstände unentbehrlich.

2.2.3 Vorgaben für flache Übergänge in der ITRS-Roadmap

In der ITRS-Roadmap [*ITRS '06*] werden die wichtigsten Anforderungen an die Fertigungstechnologie abgeleitet und zusammengefasst. Unter Berücksichtigung von physikalischen und auch technologischen Grundlagen, wie oben ausgeführt, sind dort auch Vorgaben enthalten, die die Herstellung flacher Übergänge betreffen. Zunächst sollen an dieser Stelle die allgemeinen Definitionen erläutert werden.

Eine zentrale Bedeutung besitzt die Roadmap mit der Festlegung von zu erreichenden Fertigungsdimensionen und -toleranzen, Materialeigenschaften und -konzentrationen und auch durch die Analyse von gegenwärtigen und kommenden Fertigungsmethoden, um die angestrebten Bauelementeigenschaften, wie Abmessungen und Leistungsvorgaben, aber auch Fertigungskosten usw. in den gegebenen Zeiträumen zu erreichen. Als wichtiger Leitfaden hierfür dient wiederum das Mooresche Gesetz selbst. Die Jahresvorgaben der Roadmap beziehen sich daher auch ausdrücklich auf den Beginn der Massenfertigung, bis auf einige gesonderte Planungen für die Demonstration von Prototypen sowie für die erste Einführung in die Produktion. Eine Massenfertigung erfolgt, wenn die Herstellung von 10.000 Einheiten von ICs pro Monat an einer Fertigungsstätte überschritten wird und von einem weiteren Hersteller innerhalb von drei Monaten mit der Produktion begonnen wird.

Die Entwicklung ist dabei in Technologiegenerationen (technology nodes) unterteilt, anhand der Erzielung besonderer Kennziffern durch signifikante Fortschritte in der Prozessierung, und die zugrunde liegende Fertigungstechnologie ist danach benannt. Wegen der Bedeutung dieser Kennziffern und auch um Unklarheiten zu vermeiden, wurde ab der Roadmap 2003 die jeweilige Technologiegeneration eindeutig definiert und in den Tabellen angegeben.

Die Kennziffern beziehen sich nun traditionell auf die am geringsten ausführbaren Abstände benachbarter Leiterbahnen aus Metall oder Polysilicium im Fertigungsprozess, wobei gegenwärtig noch die Produktion von Speicherbauelementen (DRAM) als Vorreiter den Maßstab bildet. Die Hälfte dieses Abstandes, der so genannte minimale Half-pitch oder DRAM feature size, ist die Kennziffer für die Technologiegeneration.

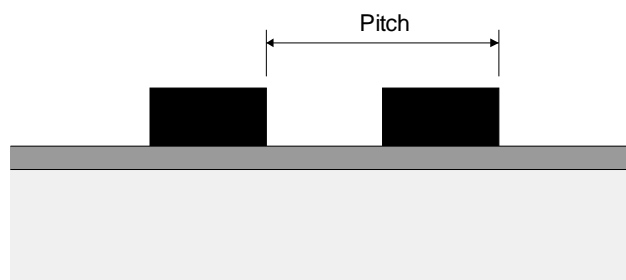


Abbildung 2.2: Schematische Schnittdarstellung einer Struktur mit zwei Leiterbahnen zur Definition des Abstandes (Pitch)

Dabei werden jedoch durchaus auch in der Fertigung anderer Bereiche geringere Dimensionen als dieses Maß innerhalb der jeweiligen Technologiegeneration benötigt und erzielt, wie z.B. die pn-Übergangstiefen oder auch laterale Dimensionen wie die Gatelänge.

Eine neue Technologiegeneration wird nun so definiert, dass ein bedeutender Technologiefortschritt mit der Reduzierung des Half-pitch auf das 0,7-fache der vorherigen Generation, bzw. auf die Hälfte gegenüber der vorletzten, einhergeht, wobei die Werte gerundet werden. Dies entspricht einem Skalierungsfaktor von ca. $\sqrt{2}$ zwischen zwei Technologiegenerationen. Die Planungen des zeitlichen Verlaufs orientieren sich an dem historischen Entwicklungstempo und den tatsächlichen Produktionseinführungen aber auch an Abschätzungen der Möglichkeiten des weiteren Wachstums. So ist traditionell von einem 3-Jahreszyklus ausgegangen worden, jedoch ist von 1998-2004 eine schnellere Entwicklung von 2 Jahren zwischen den Generationen festgestellt worden und die entsprechende Roadmap (2001) angepasst worden. In der gegenwärtigen Roadmap (2005) wird jedoch wieder mit einem 3-jährigen Zyklus für die kommende Entwicklung gerechnet.

Im oberen Teil der Abbildung 2.3 ist diese Entwicklung des Half-pitch von DRAM-Speichern dargestellt und die entsprechenden Technologiegenerationen sind bezeichnet. Zum Vergleich ist hier auch die übliche Einteilung in Generationen der Speichergröße ab einem Gigabit mit der bestehenden und erwarteten Einführung in die Massenproduktion eingezeichnet.

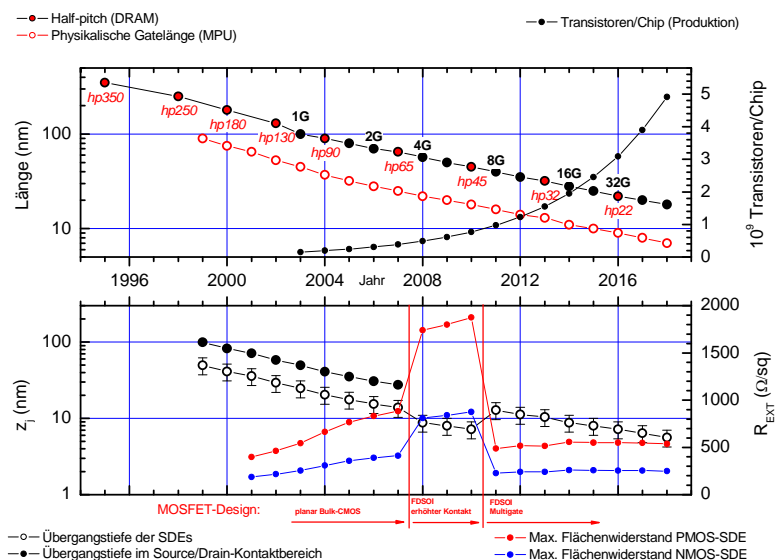


Abbildung 2.3: Vergleichende Darstellung wichtiger Kenngrößen der ITRS-Roadmap (2004) über dem Zeitpunkt der Einführung in die Massenfertigung.

Oben: Darstellung des Half-pitch von DRAM-Speichern mit abgeleiteter Bezeichnung der Technologiegeneration (DRAM-Speichergeneration zum Vergleich, wobei 1G = 1 Gigabit) im Vergleich mit der physikalischen Gatelänge von MOSFETs in Mikroprozessoren (MPU) dieser Generation und im Zusammenhang mit der Anzahl der Transistoren pro Chip (MPU).

Unten: Darstellung der Übergangstiefen der Erweiterungsbereiche von Source und Drain und entsprechende maximale Flächenwiderstände für p- oder n-Ausführung sowie Übergangstiefen im Kontaktbereich von Source und Drain unter Berücksichtigung der erwarteten vorherrschenden MOSFET-Bauform.

Weiterhin kann die parallele Entwicklung von Mikroprozessor-Bauelementen (MPU – micro processor unit) anhand der Skalierung der physikalischen Gatelänge abgelesen werden. Dieses Maß ist bei den MOSFETs dieser Bauelemente eine charakterisierende technologische Kenngröße, als physikalisch wirksamer Gatebereich, der sich nach Ende der Fertigungsschritte, insbesondere nach den selbstjustierenden Prozessen zur Gatedefinition und den Diffusionsprozessen der SDE-Gebiete, einstellt (im Gegensatz zur größeren lithographisch definierten Gatelänge). Dieses Maß ist damit noch deutlich geringer als es die nach oben

beschriebenem Abstand definierte Technologiegeneration angibt. So ist bei der 90 nm-Technologie bereits eine physikalische Gatelänge von 65 nm realisierbar. Die mit der Skalierung einhergehende zunehmende Komplexität dieser Bauelemente ist anhand der Anzahl der integrierten Transistoren je Chip verdeutlicht. Es wird erwartet, dass innerhalb der 45 nm-Technologie zwischen den Jahren 2011/12 die Zahl von 1 Milliarde Transistoren pro Chip überschritten wird.

Abgeleitet von den Skalierungsbedingungen wurde in der gegenwärtigen Roadmap die folgende Beziehung für die Festlegung der Tiefe der Übergänge im Bereich der Source-/Drain-Extensions herangezogen:

$$z_j = 0,55 L_G \pm 25\%, \quad (2)$$

wobei L_G die physikalische Gatelänge ist. Im unteren Teil der Abbildung 2.3 ist dargestellt, wie sich diese Anforderung an die Übergangstiefe im SDE-Bereich entwickelt hat und weiterhin entwickeln sollte. Auch sind die in Verbindung damit erforderlichen Flächenwiderstände für p- und n-Kanalbauelemente und die Übergangstiefen im Kontaktbereich dargestellt.

Aus dieser Abbildung ist darüber hinaus abzulesen, wie die zukünftige, voraussichtliche Einführung neuartiger MOSFET-Bauformen (siehe z.B. in [Hutchby '02; Wong '02; Huff '04; King '05]) diese Parameter beeinflussen. Das gegenwärtige Design soll ab 2007/08 durch im Kanalbereich vollständig verarmtes Silicium auf einer Isolatorschicht (SOI – silicon-on-insulator) abgelöst werden. Der Kanal und die aktiven Source-/Drainbereiche sind dann durch die Dicke der SOI-Schicht bestimmt, der Kontaktbereich von Source und Drain muss erhöht ausgeführt werden. Aufgrund günstigerer Kurzkanaleigenschaften vorrangig wegen der Vermeidung von Streukapazitäten und verbesserter Ladungsträgerbeweglichkeiten sind hier die Anforderungen an den Flächenwiderstand geringer. Mit der Einführung der Multigate-Bauweise, bei der z.B. zwei parallele Gates den Kanalbereich oberhalb als auch unterhalb begrenzen, werden die Anforderungen wieder erheblich höher sein.

Anzumerken ist an dieser Stelle, dass in Auswertung jüngster technologischer Entwicklungen eine noch aggressivere Skalierung der Übergänge vorgeschlagen wurde [Faifer '05],

$$z_j = 0,36 L_G. \quad (3)$$

Dies wird als eine Möglichkeit angesehen, in Anbetracht von Problemen bei der Einführung neuer Technologien zum Aufbau des Gate-Stapels in Verbindung mit der Nutzung neuer Materialien, wie durchgehend silicidierte Gates oder Metallgates und alternative hoch- ϵ -Gatedielektrika, die zukünftigen Entwicklungsziele im projektierten Zeitrahmen dennoch zu erreichen.

In der folgenden Abbildung 2.4 ist nun der Zusammenhang zwischen Flächenwiderstand und Übergangstiefe der SDEs dargestellt, wie er in der gegenwärtigen Roadmap gefordert wird. Zum Vergleich ist hier der unter Benutzung von Gleichung (1) berechnete Zusammenhang für eine Phosphordotierung in Silicium angegeben. Die von der Dotierkonzentration abhängigen Werte für die Beweglichkeit der Majoritätsladungsträger wurden entsprechend dem Modell von Masetti et al. [Masetti '83] ermittelt. Es wurde vom ideal abrupten Profil ausgegangen, das heißt von der für jede Kurve angegebenen konstanten Phosphorkonzentration und von vollständiger Aktivierung bis zum Übergang. Bei einer entsprechenden Oberflächenkonzentration stellen somit diese Verläufe die minimalen Werte des Flächenwiderstandes bei n-Dotierung mit Phosphor bis zur jeweiligen Tiefe dar.

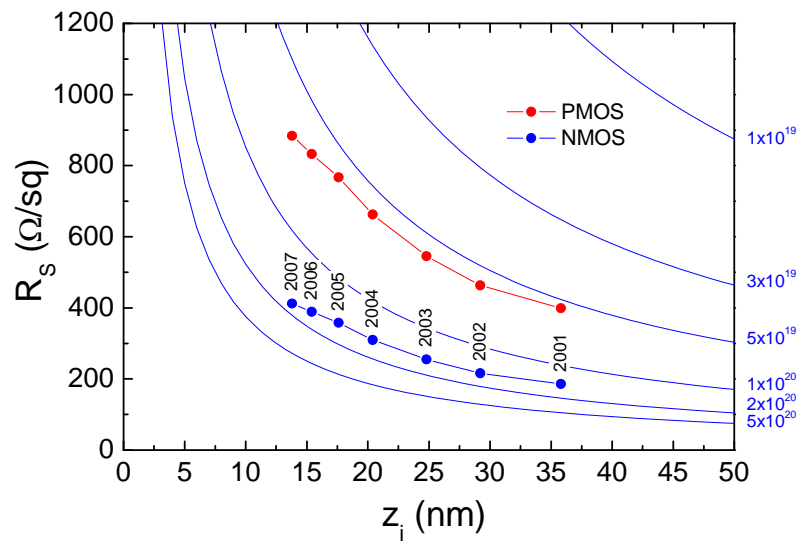


Abbildung 2.4: Geforderter Flächenwiderstand über der Übergangstiefe für Source-/Drain-Extensions planarer MOSFETs mit Jahr der Einführung in die Massenfertigung laut ITRS-Roadmap (2004).

Blaue Linien: Mittels Gleichung (1) berechneter Zusammenhang zwischen Flächenwiderstand und Übergangstiefe bei rechteckförmigem Dotierprofil mit bezeichneter Konzentration aktiver Phosphordotatoren.

Das Erreichen dieser Anforderungen ist eine technologische Herausforderung. Noch in der Roadmap von 2001 war angegeben worden, dass ab der 65 nm-Generation keine Methode bekannt wäre, die Vorgaben für die Dotierung der SDE-Bereiche einzuhalten. Mit den erst kürzlich erreichten Verbesserungen der Implantation und besonders der Annealing-Verfahren wie Laser- und Flash-Annealing erscheint es nun technologisch realisierbar, diese Anforderungen zu erfüllen [Taylor '04; Foggiato '05; Lerch '05].

Außerdem ist an dieser Stelle anzumerken, dass die Roadmap gemäß der Definition des Zeitpunktes der Einführung in die Massenproduktion gewissermaßen ein maximales Entwicklungstempo vorgibt, das nur von den führenden Herstellern eingehalten werden kann. Für andere Prozesse und Fertigungskonzepte gelten die Anforderungen für die Erzielung der geforderten Strukturen zwar entsprechend, jedoch ist der Zeitpunkt der Produktion oft später und „ältere“ Generationen sind hier noch Bestandteil der gegenwärtigen Technologie, so dass auch Entwicklungen in deren Rahmen durchaus noch von Interesse sind.

2.3 Verbreitete Herstellungsverfahren von flachen Übergängen und deren Beschränkungen

2.3.1 Ionenimplantation und thermisches Ausheilen

Die Vorbelegung mit Ionenimplantation unter Verwendung niedriger und niedrigster Beschleunigungsenergien und der anschließende Kurzeitenausheilschritt (RTA - Rapid Thermal Annealing) stellen heute die gängige und erfolgreiche Vorgehensweise für die Dotierung begrenzter Halbleiterbereiche mit planaren Topographien dar [Rubin '03].

Bei diesem Verfahren werden Atome des Dotierstoffs oder auch Moleküle, die ihn enthalten, ionisiert, so dass sie in einem elektrischen Feld beschleunigt und auf das Siliciumsubstrat gerichtet werden können. Aufgrund ihrer hohen kinetischen Energie treten sie in den Festkörper ein. Durch Energieabgabe infolge direkter Kernstöße mit Substratatomern und Anregungswechselwirkungen mit Substratelektronen kommen sie im Inneren des Festkörpers auf

Zwischengitterplätzen zur Ruhe. Durch die Anregungs- und Stoßwechselwirkungen werden Substratatome aus Gitterplätzen verdrängt, dies bedeutet eine Zerstörung der Kristallstruktur im Bereich der Reichweite dieser Wechselwirkungen im Substrat. Ein thermischer Prozess ist daher nach jeder Implantation sowohl für das Wiederherstellen (Ausheilen) der Kristallstruktur als auch für eine Aktivierung, das heißt Beförderung der implantierten Atome auf reguläre Gitterplätze, erforderlich. Nur in Bindungen mit den Festkörperatomen können Fremdatome zur Elektronenkonfiguration des Festkörpers mit Ladungsträgern wirksam beitragen.

Der Hauptvorteil dieser Methode ist die in weiten Bereichen exakte Kontrollierbarkeit des Dotierergebnisses. So sind unabhängig voneinander die Eindringtiefe (vor dem Ausheilsschritt) durch die Wahl der Ionenenergie und des Einfallswinkels der Ionen und die Dosis durch Festlegung des Ionenstroms und der Prozesszeit einstellbar, das heißt über Prozessparameter, die einer exakten Kontrolle und Regelung zugänglich sind.

Für die Herstellung sehr flacher pn-Übergänge, wie sie z.B. für die SDE-Bereiche gefordert sind, können jedoch prinzipielle Beschränkungen dieser Methode nicht umgangen werden [Agarwal '00]. Ihre Anwendung erfordert einen hohen technischen Aufwand, sowohl für den eigentlichen Implantationsschritt (ultra low energy (ULE) ion implantation), als auch für den immer erforderlichen Ausheilsschritt. Für eine Phosphordotierung finden sich in diesem Zusammenhang jedoch nur wenige Ausführungen in der Literatur [Ruffell '05]. So soll dies am Beispiel der Bordotierung, als wichtiger Gegenstand aktueller intensiver Untersuchungen, verdeutlicht werden [Privitera '02].

Trotz der sehr niedrigen Implantationsenergien unterhalb von 1 keV bei ULE-Ionenimplantation, die für flache Profile erforderlich sind, entstehen im Siliciumkristall Gitterdefekte [Moffatt '00] oberhalb der Gleichgewichtskonzentration, die während des Hochtemperaturschrittes für die elektrische Aktivierung der implantierten Spezies zu unerwünschten transienten Diffusionsphänomenen führen, wie der vorübergehend verstärkten Diffusion (TED - transient enhanced diffusion) von Bor. Letztendlich sind somit diese transienten Effekte für die Lage und den Verlauf des Überganges maßgeblich entscheidend. Die Bordiffusion in Silicium erfolgt über einen Interstitial-Mechanismus, so dass eine erhöhte Konzentration von Interstitials, die bei einer Implantation prinzipiell unvermeidlich ist, zu einer verstärkten Diffusion von Bor führt. Für flache und abrupte Borprofile ist dies sehr nachteilig. Die Verstärkung der Diffusion ist transient, da sie nur so lange besteht, wie der Nichtgleichgewichtszustand an Interstitials vorhanden ist und sie nicht z.B. an der Oberfläche vernichtet werden. Die genauen Mechanismen der TED von Bor waren und sind der Gegenstand intensiver Forschungen und sind zum Teil noch nicht endgültig aufgeklärt [Shao '03]. Dies schränkt die Möglichkeiten einer exakten Prozesssimulation und die technologische Beherrschbarkeit ein. So wird für die TED der mit sehr geringer Energie implantierten Boratome ein anderer Mechanismus vermutet als der durch das weitgehend akzeptierte TED-Modell der „Ostwald-Reifung“ von Silicium-Selbstinterstitials zu {311}-Defekten [Claverie '99]. Instabile Cluster, die Bor enthalten können, werden dafür verantwortlich gemacht, so dass die durch die Implantation erreichten geringen Eindringtiefen während der Aktivierung zunichte gemacht werden können. In zunehmendem Maße spielen dabei auch Oberflächen- und Interfaceeffekte eine maßgebliche Rolle, die in der Regel zu einer verstärkten Rekombination der Defekte führen, die Kontrollierbarkeit der Prozesse jedoch weiter erschweren. Ist daher die Implantation im sub-keV-Bereich auch durch die Verfügbarkeit technisch sehr anspruchsvoller damit aber auch kostspieliger Implantier selbst im Fertigungsmaßstab weitgehend beherrschbar, so bleibt die Aktivierung eine große Herausforderung, zu deren Durchführung verbesserte Hochtemperaturverfahren wie das Laser Annealing, bzw. Excimer Laser Annealing (ELA) Gegenstand der Untersuchungen sind [Whelan '02] oder die z.B. durch Co-Implantation von Fluor verbessert werden soll [Shaully '04].

Ein weiteres Problem ist das Channeling, d.h. das erleichterte Eindringen von implantierten Spezies innerhalb durch die Kristallstruktur bedingter Kanäle. Dieser Effekt tritt generell bei einkristallinen Materialien in Erscheinung und muss für die Dotierung flacher und begrenzter Bereiche verhindert werden, wie durch eine gezielte Amorphisierung des Oberflächenbereiches vor der Implantation (PAI - pre-amorphization implantation) [Kim '02; Li '03]. Dies stellt natürlich eine zusätzliche, nicht unerhebliche Störung der Oberfläche dar, die auch nachteiligen Einfluss auf nachfolgende Prozessschritte haben kann.

Aufgrund der ausgeprägten Vorzugsrichtung der beschleunigten Ionen bei der Ionenimplantation ist Dotierung von Flächen bei dreidimensionalen Topographien wegen deren Abschattungseffekten nur höchst unzureichend realisierbar. So konnten bei einer Implantation von Bor unter einem flachen Einfallswinkel von 10° nur 40 % der Ionen in die vertikale Seitenwand, auf die der Strahl auftrifft, auch tatsächlich eindringen, während 10% der Ionen in der gegenüberliegenden Wand zu finden waren [Hattori '01].

Neuere Entwicklungen von Leistungsbau-elementen, wie den CoolMOS [Lorenz '99] oder den RESURF DMOS [Glenn '00], zeigen den Bedarf der Dotierung von nichtplanaren Strukturen. Diese neuen Konzepte weisen im Driftbereich parallele alternierend hoch p- und n-dotierte Bereiche auf, um im Aus-Zustand Ladungsträgerkompensation für hohe Sperrspannungen (z.B. etwa 800 V) bei gleichzeitig äußerst niedrigen Vorwärtsspannungsabfällen zu erreichen. Abhängig vom speziellen Anwendungsbereich und Bauelementedesign können dann aber diese „Superjunctions“ nicht mehr durch Diffusion beider Dotierstofftypen von der Substratoberfläche aus erzeugt werden, sondern es wird die Dotierung in Trench-Strukturen zur Herstellung der Superjunctions erforderlich [Hattori '01].

Bei hohem Aspektverhältnis von Trenchstrukturen kann prinzipiell durch die Ionenimplantation diese Aufgabenstellung nicht mehr gelöst werden und neue technologische Ansätze mit ausreichend hoher Homogenität bei dreidimensionaler Topographie, wie die Dotierung aus der Gasphase, sind erforderlich.

2.3.2 Alternative Dotierverfahren

Aufgrund der Beschränkungen und Nachteile der konventionellen Ionenimplantation wurden alternative Dotierverfahren wie Infusion Doping [Skinner '06], PGILD (projection-gas-immersion laser doping), PIII (plasma immersion ion implantation) und im folgenden vorgestellte Methoden entwickelt und untersucht (siehe z.B. [Jones '98] und Zitate darin). Nachfolgend soll nur auf solche Verfahren kritisch eingegangen werden, die ohne Ionenbeschuss der Oberfläche und damit weitgehend strukturdefektfrei durchgeführt werden können (non-radiation doping), entsprechend der Methode, die als Grundlage der Untersuchungen in der vorliegenden Arbeit gewählt wurde.

2.3.2.1 Rapid Thermal Doping und Solid Phase Diffusion

Seit 1987 wird von RTD-Prozessen (RTD - rapid thermal doping) unter Benutzung einer Feststoffquelle berichtet [Kim '87; Kim '90]. Dieser RTD-Prozess mit fester Dotierstoffquelle, auch Solid Phase Diffusion (SPD) genannt [Saito '93], läuft als Zweistufenprozess ab. Nach der Vorbelegung bei mittleren Temperaturen wird die Dotierstoffquelle entfernt und der Dopant wird in einem zweiten Temperaturschritt unter Inertgas eingetrieben [Cho '92]. Als Vorbelegung wurden Nitride [Kim '87], dotierte Oxide [Souza '88; Thakur '91; Usami '92; Ono '95] oder Polymerfilme [Usami '92] untersucht, oder auch dotierte epitaktische Schichten aus Si oder SiGe [Nakahara '96; Uchino '01]. Neben der Epitaxie für die Abscheidung der Vorbelegungsschicht wurden dabei konventionelle CVD-Verfahren (chemical vapor depo-

sition) [Souza '88; Thakur '91] oder das Aufschleudern in Spin-On-Verfahren (SOD - spin-on doping) [Usami '92] eingesetzt. Zum Teil wurde, wie z.B. in [Souza '88] beschrieben, die Ausdiffusion des Dotierstoffes im anschließenden RTA-Schritt durch das Aufbringen einer weiteren, undotierten CVD-SiO₂-Deckschicht begrenzt. Allen diesen Vorgehensweisen ist gemeinsam, dass die Vorbelegung der Oberfläche mit Dotierstoff in einem zusätzlichen Prozessschritt vor dem Eintreibschritt erfolgt.

In einigen vorgestellten Arbeiten [Gopalan '03; Oh '04] wurde der Einsatz des SOD-Verfahrens für n-Dotierung in Prozessfolgen zur Herstellung von sub-100 nm Bauelementen beschrieben. Dabei wurden z.B. auf SOI-Substraten mit Arsendotierung flache Übergänge¹ von 27 nm mit Flächenwiderständen von 260 Ω/sq erzielt [Oh '04].

Fraglich ist, wie genau die Dosiskontrolle und Reproduzierbarkeit bei den Spin-On-Verfahren gestaltet werden kann, die letztendlich von der Qualität und Homogenität der aufgeschleuderten Schichten abhängen. Auch die Durchführbarkeit des Dotierens von dreidimensionalen Topographien, insbesondere von Trenches, erscheint ungewiss.

2.3.2.2 Vapor Phase Doping

Einige Arbeiten befassen sich mit der Gasphasendotierung (VPD - vapor phase doping oder RVD - rapid vapor phase doping). So wurden z.B. Untersuchungen über VPD von Bor [Kiyota '93; Kiyota '96; Theunissen '99], Arsen [Ransom '94] und auch Phosphor [Kiyota '94; Sato '98] veröffentlicht. Diesen Arbeiten ist gemein, dass die Siliciumoberfläche bei hohen Temperaturen dem Dotiergas unter Verwendung eines Trägergases ausgesetzt wurde, so dass der Dotierstoff aus der Gasphase über einen Adsorptions-/Dissoziationsvorgang auf der Substratoberfläche direkt in den Halbleiter diffundieren konnte. Der Vorteil dieser Verfahren ist der Transport des Dotierstoffes in der Gasphase ohne Vorzugsrichtung der Moleküle (von Strömungsvorgängen abgesehen), so dass nur ein vergleichsweise geringer apparativer Aufwand nötig ist und prinzipiell auch Oberflächen mit dreidimensionalen Topographien dotierbar sind. Eine energetische Anregung des Gases erfolgt bei konventionellen VPD-Verfahren nicht, so dass Strukturschäden durch Stoßprozesse wie bei der Implantation nicht auftreten können.

Es wurde gezeigt, dass mit diesem Verfahren Bordotierungen mit Übergangstiefen 40 nm und Flächenwiderständen von 400 Ω/sq in einem CMOS-kompatiblen Prozess für die Dotierung der SDE-Bereiche von p-MOSFETs möglich sind [Uchino '04]. Für Arsen- bzw. Phosphordotierungen liegen die bisher mit diesem Verfahren erzielten Ergebnisse nicht im Bereich der für ultraflache Profile geforderten Werte, daher musste auch für den in [Uchino '04] beschriebenen Prozess für die Dotierung der n-MOSFET-SDE-Bereiche auf ein SPD-Verfahren unter Anwendung von Phosphorsilikatglas als Dotierstoffquelle zurückgegriffen werden.

Die maximal erreichbare Dosis bei VPD-Verfahren ist begrenzt durch die Festkörperlöslichkeit und Desorptionsprozesse in die Gasphase; die veröffentlichten Ergebnisse für Phosphor und Arsen lassen die Dotierstoffkonzentration zu gering erscheinen für die geforderten Flächenwiderstände. Nachteilig bei den VPD-Verfahren ist außerdem eine nur ungenügende Dosiskontrolle und damit auch Homogenität der Dotierung über den gesamten Wafer. Die Dosis ist kritisch abhängig von der Konzentration des Dotierstoffes in der Gasphase, insbesondere im Bereich der Oberfläche des Substrates, und in der Regel auch von der Oberfläche selbst. Konvektionserscheinungen des zumeist unter Atmosphärendruck

¹ Wie bereits erwähnt, ist hier zu beachten, dass sich die Übergangstiefen in der Regel nicht auf die tatsächlichen Tiefen der metallurgischen pn-Übergänge beziehen, die in den entsprechenden Experimenten, oft auf Testsubstraten, erzielt wurden. Vielmehr wird zumeist der Übergang so *definiert*, dass damit den realen Verhältnissen der Bauelemente Rechnung getragen wird. So können die in der Literatur angegebenen Übergangstiefen auf unterschiedliche Konzentrationen bezogen sein.

strömenden Gases und Reaktionen in der Gasphase und an Wänden des Reaktors erschweren einen kontrollierten reproduzierbaren Prozessablauf. Außerdem kann der Profilverlauf kaum beeinflusst werden. Der annähernd exponentielle Abfall der Dotierstoffkonzentration im Halbleiter bei konventionellen VPD-Verfahren ist für die geforderten Eigenschaften der pn-Übergänge ungünstig.

2.3.2.3 Atomic Layer Doping

Mittels neuerer mehrstufiger Prozesse wurde versucht, Nachteile des VPD-Verfahrens zu vermindern. Tillack [Tillack '98] stellte ein so genanntes Atomic Layer Doping (ALD) für die Dotierung von SiGe mit Bor bzw. Phosphor vor. Bei diesem Verfahren erfolgt die Adsorption des Dotierstoffes in einem von der Diffusion getrennten Prozessschritt, wie es typisch ist für die so genannten atomar kontrollierten Prozesse. Die Dotierung, in [Tillack '00] für den Basisbereich von Heterojunction-Bipolartransistoren, erfolgte durch dissoziative Adsorption von Phosphin bzw. Diboran auf SiGe-Filmen, die durch CVD abgeschieden wurden. Dabei zeigte sich, dass bei der Phosphoradsorption eine Sättigung der P-Belegung bei weniger als einer Monolage auftritt, bedingt durch die besonderen Bindungsverhältnisse von 5-wertigen Elementen mit der Si-Oberfläche [Miotto '01b], die zu einer Art Terminierung der Oberfläche führt. Bei der Adsorption von Bor wurde keine Sättigung beobachtet [Tillack '00], was damit die Prozesskontrolle der Bordotierung mit diesem Verfahren erschwert. Eine Erweiterung dieses Prozessansatzes zu einem mehrfachen ALD (multiple ALD) für Phosphor wurde in [Shimamune '04] für eine P-Dotierung während einer Siliciumepitaxie demonstriert, durch den wiederholten Wechsel von silan- und phosphinhaltiger Prozessatmosphäre.

Kim et al. [Kim '00] stellten eine Dotiertechnologie für die Bordotierung ultra-flacher Übergänge vor. Hier wurde auch eine zusätzliche Oxidschicht zur Verminderung der Ausdiffusion aufgebracht, was zu einem 3-Schrittprozess führt: eine Boradsorption aus Diboran, ein Aufwachsen einer 30 nm CVD-SiO₂-Deckschicht bei einer Temperatur von 450 °C und dem Eintreibeschritt in Argon-Atmosphäre bei 950 °C - 1050 °C für 5 s. Damit wurden sehr flache pn-Übergänge < 25 nm mit Flächenwiderständen im Bereich von 2,5 kΩ/sq erzielt und ihre Anwendung in SDE-Bereichen von MOSFETs demonstriert. Das nichtbeschränkte Wachstum der Boradsorbatschicht musste über eine genaue Kontrolle der Diborandosis erfolgen. Nachteilig ist weiterhin, dass für den zusätzlichen Decklayer-Depositionsschritt eine separate Prozesskammer erforderlich ist.

In einem anlogem Prozess für eine Arsendotierung mit konventionellem RTA erreichten Song et al. [Song '00] ultraflache Übergänge von 20 nm Tiefe mit Flächenwiderständen unterhalb von 2 kΩ/sq unter Verwendung einer 30 nm APCVD-SiO₂-Deckschicht. Koh et al. [Koh '02] beschreiben solch ein Verfahren für die Arsen-Dotierung von SOI-MOSFETs mit Hinblick auf den Einsatz eines Laser-Annealing für die Diffusion. Hier erfolgte eine Oberflächenadsorption von Arsen aus Arsin, ein Aufwachsen einer 130 nm CVD-Schicht SiO₂ und der Eintreibeschritt durch Laser-Annealing bzw. konventionelles RTA bei 950 °C - 1080 °C für 3 s. Mit dem RTA-Verfahren wurden dabei Flächenwiderstände von 1,2 kΩ/sq bei Übergangstiefen von 23 nm erreicht.

Die Anwendung der in [Kim '00; Koh '02] vorgestellten Verfahren für die Herstellung von Flash Speichern mit an Seitenwand-Bereichen von SOI-Strukturen angeordnetem Kanal und Floating Gate wurde in [Choi '03] beschrieben. Dies ist ein Beispiel für die erfolgreiche Integration der ALD-Verfahren für dreidimensionale Topographien in einem CMOS-Prozess. Dabei wurde Bor-ALD für die Dotierung des oberen Bereichs der SOI-Oberfläche zur Verhinderung parasitärer Kanäle und Arsen-ALD für die SDE-Bereiche der Seitenflächen der SOI-Region eingesetzt. Insbesondere solch eine Seitenwanddotierung kann mit einer konventionellen Ionenimplantation nicht durchgeführt werden.

Nachteilig bei diesen Verfahren ist die Schwierigkeit der genauen Dosiskontrolle bei der nicht selbstbeschränkten Borbelegung bzw. des Erzielens von Belegungen von Arsen bzw. Phosphor, die über eine Monolage hinausgehen sollen. Außerdem ist ein definierter Zustand der Oberfläche entscheidend für eine kontrollierte Adsorption.

Hervorzuheben ist jedoch, dass bei den Diffusionsverfahren für die Dotierung neben dem Bereitstellen der Dotierstoffquelle immer ein Hochtemperaturschritt wie auch bei der Implantation erforderlich ist, der kritisch für das Dotierergebnis ist, sowohl bezüglich der Diffusionstiefe und der Dotierhomogenität. Die Erzielung von kürzesten Prozesszeiten und homogenen und reproduzierbaren Temperaturverteilungen stehen dabei im Vordergrund.

2.4 Wahl des Dotierverfahrens

Für die hier beschriebenen Untersuchungen zur Siliciumdotierung wurde ein Verfahren angestrebt, das ohne Implantation durchgeführt werden kann. Es sollte eine einfache Dotierung aus der Gasphase heraus gestatten. Für die Experimente wurde Phosphor aus einer Phosphinquelle ausgewählt. An dieser Stelle sollen die wichtigsten Gründe zusammengefasst werden, die in die Überlegungen zur Wahl des Verfahrens eingeflossen sind.

Alternative zur Ionenimplantation mit einfachem Verfahren

Zur Verfügung stand eine gewöhnliche Kurzzeitprozessanlage für Wafer mit 150 mm Durchmesser. Entsprechende Anlagen sind in jeder CMOS-Fertigung für kurze Hochtemperaturschritte vorhanden, so z.B. für das Ausheilen nach Implantationen oder für Kurzzeit-Beschichtungsverfahren. Eine CMOS-kompatible Integration von Prozessen ist damit leicht möglich.

Die Anlage ist vakuumtauglich, so dass Untersuchungen in einer reinen kontrollierten Niederdruck-Prozessatmosphäre durchgeführt werden konnten, mit dem Ziel eines möglichst geringen Gaseinsatzes. Dies bedeutet geringere Kosten für Verbrauchsmaterial und auch geringere Umweltbelastung im Vergleich zu Prozessen unter Atmosphärendruck.

Transportprozesse aus der Gasphase ohne Beschleunigung und ohne eine ausgeprägte Vorzugsrichtung von Gasteilchen vermeiden das Entstehen von Defekten der Kristallstruktur und erlauben prinzipiell die Dotierung von nichtplanaren Strukturen mit dreidimensionalen Topographien, wie tiefe Gräben oder Wände.

Eine Diffusion aus einer erschöpflichen Quelle an der Oberfläche ist gegenüber einer Diffusion direkt aus der Gasphase von Vorteil, da damit ein gaußförmiger Profilverlauf mit günstigeren elektrischen Eigenschaften der Übergänge zu erwarten ist.

Phosphordotierung aus Phosphinquelle

Für einen oberflächenkontrollierten Prozess erweisen sich Elemente aus der 5. Hauptgruppe günstig, da hier eine selbstbeschränkte Anlagerung erwartet wird. Für die Experimente wurde daher die n-Dotierung bevorzugt.

Für die Herstellung von flachsten Übergängen sollten die Elemente Arsen oder Antimon günstiger im Vergleich zu Phosphor sein, aufgrund der geringeren Diffusionskoeffizienten dieser Elemente in Silicium. Für die Experimente wurde dennoch Phosphor favorisiert, da es

unkritischer ist im Vergleich zu Arsen, insbesondere beim Entfernen von eventuellen Kontaminationen in der Anlage, und auch günstiger in der Beschaffung.

Es existiert bereits eine große Anzahl von grundlegenden Untersuchungen über die Anlagerung von gasförmigem Phosphin bzw. Phosphor an Siliciumoberflächen. Über die Anwendung von solchen Belegungen als Dotierstoffquelle für Silicium wurde hingegen bisher wenig berichtet.

Grundlegende Ergebnisse einer Phosphordotierung aus einer PH_3 -Quelle sollten auf Dotierungen aus AsH_3 oder SbH_3 übertragbar sein.

2.5 Zusammenfassung

Die Bildung flachster pn-Übergänge, besonders im Bereich der Source-/Drain-Extensions von MOSFETs ist einer der wichtigen Prozesse in der modernen Halbleitertechnologie für integrierte Schaltungen und wird in der ITRS-Roadmap als Herausforderung bezeichnet. Solche flachen sehr hoch dotierten Bereiche sind erforderlich, um unerwünschte Kurzkanaleffekte unterdrücken zu können, bei gleichzeitiger Minimierung parasitärer Widerstände. Der Serienwiderstand von Source nach Drain muss für die Erzielung hoher Source-Drain-Ströme bei niedriger Leistungsaufnahme gering gehalten werden. So wird z.B. bereits für die Herstellung von MOSFETs der 65 nm-ULSI-Technology (Produktionsjahr 2007), wo physikalische Gate-Längen von 25 nm auftreten, für die SDE-Bereiche gefordert, die Tiefe von abrupten pn-Übergängen nach Aktivierung auf 13,8 nm zu beschränken, bei einem maximalen Flächenwiderstand von 884 Ohm/sq für p-MOS- und von 412 Ohm/sq für n-MOS-SDEs.

Die erwarteten Anforderungen für die kommenden Technologien, wie sie in der Roadmap dargelegt sind, erfordern Fertigungsverfahren, die zum Teil bereits verfügbar sind, zumeist jedoch noch nachhaltig der Entwicklung und Optimierung bedürfen, insbesondere auch bezüglich laufender Kosten und Investitionskosten.

Als eine Alternative zur überwiegend eingesetzten Ionenimplantation, im Hinblick auf deren Beschränkungen bezüglich Defektbildung und Bearbeitung dreidimensionaler Strukturen, sollte die Dotierung aus der Gasphase zur Schaffung von oberflächennahen pn-Übergängen in Silicium untersucht werden.

3. Grundlagen zum Dotierprozess mit Phosphor aus einer Oberflächenquelle

Phosphor wirkt als Elektronendonator im Silicium. Als billiges und technologisch leicht verfügbares Material wird es oft für die Dotierung eingesetzt. Aufgrund seiner Bedeutung als Dotierstoff und als Grundlage nachfolgender Untersuchungen soll es näher beschrieben werden. Die gasförmige Verbindung Phosphin spielt als Ausgangsstoff für Dotierprozesse wegen der Vorteile einer guten Dosierbarkeit und Prozesskontrolle eine noch bedeutendere Rolle als fester Phosphor. Die Kenntnis seiner Eigenschaften ist dennoch wichtig für das Verständnis und die Beurteilung von Abscheidungsprozessen.

Wirksam wird der Dotierstoff Phosphor im Silicium erst nach Einbau in den Kristallverband. Die Diffusion im Festkörper ist dabei der zentrale Transportvorgang für die Ausbreitung und Aktivierung des Dotierstoffs. So soll ein Überblick über die Diffusionsmechanismen von Dotierstoffen in Silicium gegeben werden, die wichtig sind für das Verständnis der allgemeinen Theorie der Diffusion. Durch das Verstehen und Modellieren der Vorgänge bei der Dotierung, sollen Methoden und Verfahren für deren technologische Beherrschung gefunden und optimiert werden.

3.1 Phosphor für die Dotierung von Silicium

3.1.1 Eigenschaften von Phosphor

Phosphor ist ein chemisches Element der 5. Hauptgruppe, eingebunden in das Siliciumkristall wirkt er somit als Elektronendonator. Er steht in der Elementhäufigkeit der Litho- und Biosphäre mit 0,1 Gewichtsprozent an der 13. Stelle. Man findet Phosphor in der Natur nie in der reinen Form sondern nur in Form von Derivaten der Phosphorsäure H_3PO_4 , ausschließlich als Isotop $^{31}_{15}\text{P}$. Der Name leitet sich vom griechischen Wort „phosphorus“ ab, was „Lichtträger“ bedeutet. Er bezieht sich auf den Leuchteffekt¹ von weißem Phosphor, der bereits bei der ersten Darstellung aus Harn im Jahre 1669 durch Henning Brand beobachtet wurde. Eine gute Übersicht über die Eigenschaften von Phosphor, wie sie hier kurz beschrieben werden sollen, findet sich in z.B. in [Holleman '95].

Fester Phosphor kommt in kristallinen Modifikationen als weißer, violetter und schwarzer und in amorpher Form als roter Phosphor vor. Der weiße Phosphor ist Produkt einer technischen Gewinnung des reinen Elements aus Phosphaterzen (Apatite, Calciumphosphat). Er ist aus P_4 -Tetraedermolekülen, die durch jeweils drei Einfachbindungen zwischen vier Phosphoratomen gebildet sind, in kubischer Gitteranordnung kristallisiert. Mehrfachbindungen treten im Festkörper nicht auf, erst oberhalb von 800 °C bilden sich P_2 -Moleküle. Die thermische Stabilität der Festkörpermodifikationen nimmt bei Raumtemperatur von weißem über roten und violetten zu schwarzem hin zu, jedoch ist hier die Umwandlungsgeschwindigkeit so gering, dass neben dem bei Raumtemperatur und Atmosphärendruck eigentlich stabilen schwarzen Phosphor auch alle anderen Formen als metastabile Modifikationen existieren.

¹ Das im Dunklen sichtbare bläuliche Leuchten als eine Form der Chemolumineszenz beruht auf einer Oxidation in Luftsauerstoff von sublimiertem weißem Phosphor zu P_2O_3 , das sich dann statt unter weiterer Wärmeentwicklung unter Energieabgabe in Form von Licht zu P_2O_5 umwandelt.

Unter Atmosphärendruck schmilzt weißer Phosphor bei 44,25 °C zu einer farblosen Flüssigkeit, und er siedet bei 280,5 °C unter Bildung eines farblosen Dampfes. Beim Tempern bildet sich aus der flüssigen Phase oberhalb von 200 °C bzw. aus der gasförmigen Phase zwischen 450 und 550 °C fester roter Phosphor, der erst bei Temperaturen um die 600 °C unter Bildung von P₄-Molekülen sublimiert oder durch weiteres langes Erwärmen oberhalb von 550 °C in violetten Phosphor umgewandelt werden kann. Schwarzer, halbleitender Phosphor kann unter sehr hohem Druck hergestellt werden, ist aber auch nur bis 550 °C stabil, wonach er sich in violetten Phosphor umwandelt. Die Sublimationstemperatur von violetter Phosphor beträgt 620 °C. Im Gegensatz zu weißem Phosphor bestehen die anderen festen Modifikationen nicht aus P₄-Molekülen, sondern aus komplizierter gebauten Schichtgittern, bzw. amorphen Polymerstrukturen im Falle des roten Phosphors.

Die chemischen und physiologischen Eigenschaften des reinen Phosphors sind stark von der Modifikation abhängig. Weißer Phosphor ist aufgrund seiner Struktur äußerst reaktionsfähig. (Hierzu zählt auch die flüssige und gasförmige Form, die aus P₄-Molekülen besteht.) In fein verteilter Form ist er an der Luft selbstentzündlich, in kompakter Form verbrennt er ab ca. 50 °C zu Phosphorpentoxid (P₂O₅ = ½ P₄O₁₀),



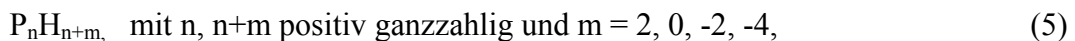
Wegen der starken Affinität zu Sauerstoff ist er ein sehr starkes Reduktionsmittel. Weißer Phosphor ist sehr giftig, der MAK-Wert liegt bei nur 0,1 mg/m³. Er kann auch durch die Haut resorbiert werden, größere Mengen von P₄ auf der Haut verursachen jedoch schwere Verbrennungen mit tiefen, schlecht heilenden Wunden. Chronische Vergiftungen führen zu Knochendegeneration.

Roter Phosphor ist hingegen ungiftig. Er ist weniger reaktionsfreudig als weißer Phosphor, leuchtet nicht und entzündet sich erst ab 300-400 °C, explodiert aber im Gemisch mit starken Oxidationsmitteln. Noch geringere Reaktionsfähigkeiten zeigen die violette und schwarze Form.

3.1.2 Eigenschaften von Phosphin

Eine Wasserstoffverbindung von Phosphor ist das Monophosphan PH₃. Diese Verbindung wird vor allem in technischem Zusammenhang oft noch Phosphin genannt. In der vorliegenden Arbeit wurde auch diese Bezeichnung verwendet.

Der dreibindige Phosphor bildet eine große Zahl von Wasserstoffverbindungen mit der allgemeinen Summenformel



aus. Verbindungen dieser Gruppe enthalten Phosphor mit jeweils drei einfachen σ-Bindungen, daher werden diese Verbindungen chemisch auch als gesättigte Phosphorwasserstoffe oder Phosphane bezeichnet.

Die Struktur von PH₃ ist pyramidal mit dem Phosphor an der Spitze. Der Bindungswinkel von H-P-H beträgt darin 93,6°.

Die technische Darstellung erfolgt durch die Umsetzung von weißem Phosphor entweder bei erhöhter Temperatur mit Natronlauge,



oder oberhalb von 250 °C mit Phosphorsäure und Wasser im Autoklaven,



Entstehende Verunreinigungen mit Diphosphan P_2H_4 lassen sich durch fraktionierende Destillation und Kondensation vom PH_3 trennen.

Phosphin ist ein farbloses, giftiges, nach Knoblauch oder verfaultem Fisch riechendes Gas. Der Umgang mit diesem Stoff ist nicht unproblematisch. Die Sicherheitsanweisungen sind unbedingt zu beachten. Der MAK-Wert beträgt nur 0,15 mg/m³ bzw. 0,1 ppm, jedoch erst ab Konzentrationen von 2 ppm ist es wahrnehmbar. Der Toxizitätswert LD₅₀ beträgt 20 vpm. Vergiftungen treten nur durch Phosphinaufnahme über die Atemwege ein. Es ist ein allgemein wirkender Giftstoff für Nervensystem, Blut und Stoffwechsel. Eine Konzentration von 2000 ppm verursacht den Tod nach kurzer Zeit, Konzentrationen von 290-600 ppm für die Dauer von 6 - 60 min sind für den Menschen lebensgefährlich. Bei geringen Konzentrationen können Vergiftungssymptome nach einer Latenzzeit von mehreren Stunden auftreten. Die Aufnahme von Phosphin führt zu Blutdruckabfall, Erbrechen, Krämpfen, Lungenödemen und Koma.

Bei Atmosphärendruck wird das Gas ab -87,74 °C flüssig und erstarrt bei -133,78 °C. Es ist unter Normalbedingungen nur wenig in Wasser löslich (330 mg/l), die Lösung reagiert neutral. Mit einer Dichte von 1,45 kg/m³ ist das Gas schwerer als Luft, die Molmasse beträgt 33,997 g/mol.

Bei erhöhter Temperatur, oberhalb von 375 °C, zersetzt sich Phosphin zu Wasserstoff und Phosphor, ein rascher Zerfall erfolgt ab 550 °C. Reines Gas entzündet sich an Luft bereits bei 150 °C unter Verbrennung zu Phosphorsäure,



Gas-Luftgemische sind explosionsfähig, bereits Spuren von Diphosphan oder anderen Verunreinigungen bewirken eine Selbstentzündung an der Luft.

3.1.3 Anlagerung von Phosphin an der reinen Siliciumoberfläche

Über das Verhalten von Phosphin an Siliciumoberflächen existieren zahlreiche Veröffentlichungen. So sei an dieser Stelle z.B. auf [*Miotto '01b*] und [*Miotto '01a*] verwiesen und die Zitate darin. Es wurde berichtet, das Phosphin leicht an (100)-Siliciumoberflächen adsorbiert, eine nichtdissoziative Anlagerung von Phosphin an den freien Bindungen der (2x1)-rekonstruierten Oberfläche wurde bei Raumtemperatur beobachtet. Mit steigenden Temperaturen findet eine Dissoziation von Phosphin an der Oberfläche statt, diese ist abhängig von der anfänglichen PH_3 -Bedeckung und damit von der Anzahl der verfügbaren freien Bindungen, die als Bindungsstelle für den bei der Dissoziation freigesetzten Wasserstoff dienen. Eine Wasserstoffdesorption von H-terminierten Bindungen der Siliciumoberfläche wird ab Temperaturen von 400 °C beobachtet, was weitere Bindungsstellen für die Adsorption von PH_3 bzw. P bereitstellt.

Der Zustand der Adsorption von vollständig dissoziiertem Phosphin ist energetisch günstiger als das Ausgangssystem aus PH_3 -Gas und der freien Siliciumoberfläche. Eine Monolage

Phosphor bei vollständiger Bedeckung besteht aus P-P-Dimeren und entspricht einer Flächenkonzentration von $6,78 \cdot 10^{14} \text{ cm}^{-2}$, der Dichte der ungesättigten Bindungen einer freien (100)-Siliciumoberfläche.¹

Die Phosphorbelegung terminiert die Siliciumoberfläche chemisch und elektrisch [Bozso '91], [Srivastava '02], da der 5-wertige Phosphor ein nichtbindendes Elektronenpaar behält. Dies führt zu einer Sättigung der Phosphorbedeckung, wie es auch von Yu und Meyerson [Yu '84] beobachtet wurde. Eine maximale Bedeckung wurde dort bei 550 °C anhand von thermischen Desorptionmessungen ermittelt.

Eine weitere Erhöhung der Temperatur führt zu einer Desorption von Phosphor von der Oberfläche. Dies ist insbesondere auch bei den Temperaturen der Fall, bei denen eine signifikante Diffusion von Phosphor in Silicium auftritt.

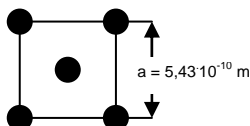
3.1.4 Festkörperlöslichkeit von Phosphor in Silicium

Um als Dotierstoff wirken zu können, muss dieser von der Oberfläche des Festkörpers auf Gitterplätze in seinem Inneren gelangen. Ein Maß für diese Aufnahmefähigkeit an Fremdstoffen stellt die Festkörperlöslichkeit dar, sie ist insbesondere bei Hochtemperaturprozessen, die unter Gleichgewichtsbedingungen ablaufen, wie die konventionellen Wannen-Dotierungen im Batch-Prozess aus der Gasphase, die entscheidende Größe, die den Fremdstoffeintrag bestimmt.

Die Festkörperlöslichkeit ist die maximale Konzentration des Fremdmaterials, die unter Gleichgewichtsbedingungen im Festkörper enthalten oder „gelöst“ sein kann, ohne eine neue Phase zu bilden. In der Praxis ist die Angabe der Festkörperlöslichkeit nur sinnvoll für höhere Temperaturen, wenn sich durch Diffusion in zeitlich vertretbarem Rahmen ein Gleichgewicht der Konzentrationen und ggf. Phasen auch einstellen kann. Sonst, insbesondere auch beim schnellen Abkühlen von höheren Temperaturen, bei denen eine höhere Festkörperlöslichkeit auftreten kann, kommt es zum so genannten Quenching, d.h. zum Einfrieren von Instabilitäten. So kann dann eine Konzentration von Fremdstoff oberhalb der Löslichkeit vorliegen, ohne dass sich eine neue Phase oder Ausscheidung (Präzipitat) ausbildet, da die Fremdstoffe bei der geringen Temperatur zu immobil sind, um den Gleichgewichtszustand auszubilden. Dies ist insbesondere bei hohen Dotierungen in Silicium bei Raumtemperatur der Fall. Im Silicium besitzen viele Stoffe eine retrograde Löslichkeit, mit dessen Maximum kurz unterhalb des Schmelzpunktes. Dabei ist gemäß der Definition einer Festkörperlöslichkeit diese in der Schmelze gleich null.

Entscheidend für die Funktion elektrischer Bauelemente ist für Stoffe, die der Dotierung dienen, die Konzentration elektrisch aktiver Fremdatome. Sie ist auch abhängig von der Art der Aufnahme der Fremdatome in den Festkörper, d.h. ob und in welcher Konfiguration Gitterplätze durch diese eingenommen wurden, und sie ist auch begrenzt durch kinetische Effekte, da die Gleichgewichtszustände nicht instantan erreicht werden können. Diese aktive Konzentration ist somit in der Regel geringer als die chemische Konzentration. So genügt es in der Regel nicht, allein z.B. durch SIMS die chemische Konzentration zu bestimmen, um

¹ Bei einer (2x1)-rekonstruierten Si-(100)-Oberfläche besteht pro Oberflächenatom eine freie ungesättigte Bindung. Die Flächendichte der Oberflächenatome kann aus dem Si-Atomabstand im Kristall bestimmt werden:



In dieser Elementarflächeneinheit der unrekonstruierten Oberfläche befinden sich zwei Atome pro Zelle, das heißt, $2 \text{ Atome} / a^2 = 6,78 \cdot 10^{14} \text{ cm}^{-2}$.

den Dotiererfolg zu bewerten, da hierbei auch inaktive Fremdatome aus Präzipitaten, Zwischengitterplätzen und neutralen Clustern zum Messergebnis beitragen.

Für die Dotiertechnologie von Silicium mit Phosphor sind Kenntnisse der Gleichgewichtseigenschaften des binären Systems Si-P erforderlich. Aus dem durch das Studium von SiP-Verbindungen gewonnenen Phasendiagramm ergibt sich, dass sich unterhalb der eutektischen Temperatur von 1131 °C unter Normaldruck die Lösung von Phosphor in Silicium im Gleichgewicht mit orthorhombischem SiP befindet [Olesinsky '86]. Für die Dotierprozesse jedoch von besonderer Bedeutung ist das Vorhandensein von elektrisch aktivem Phosphor. Es wurde beobachtet, dass die Ladungsträgerdichte in getempertem Silicium mit der Konzentration von gelöstem Phosphor ansteigt, dafür jedoch eine temperaturabhängige obere Grenze in der Größenordnung von einem Atomprozent existiert. Dabei kann hingegen die chemische Phosphorkonzentration Größenordnungen von zehn Atomprozent erreichen, ohne dass eine ausgefallte Phase beobachtet wird; das heißt, dass hierbei auch der inaktive Phosphor substitutionelle Gitterpositionen einnimmt. So wird angenommen, dass, wie auch für die Bildung von Arsenclustern beschrieben, komplexe Punktdefekte (E-Center) entstehen. Aus Experimenten mit Silicium, das über die Sättigungsgrenze hinaus mit Phosphor mittels Ionenimplantation dotiert wurde, konnten Solmi et. al. [Solmi '96] die Temperaturabhängigkeit der Festkörperlöslichkeit, d.h. der Sättigungskonzentration von Phosphor im Gleichgewicht mit SiP-Präzipitaten, bestimmen:

$$C_{sat} = 2,45 \cdot 10^{23} e^{-\frac{0,62}{kT}} \text{ cm}^{-3}. \quad (9)$$

Dies wurde mit der Gleichgewichtskonzentration des elektrisch aktiven Phosphors verglichen, die durch die Ladungsträgerkonzentration, deren Temperaturabhängigkeit gemäß

$$n_e = 1,3 \cdot 10^{22} e^{-\frac{0,37}{kT}} \text{ cm}^{-3} \quad (10)$$

ermittelt wurde, bestimmt ist. In der Abbildung 3.1 sind diese Konzentrationen in Abhängigkeit von der Temperatur dargestellt.

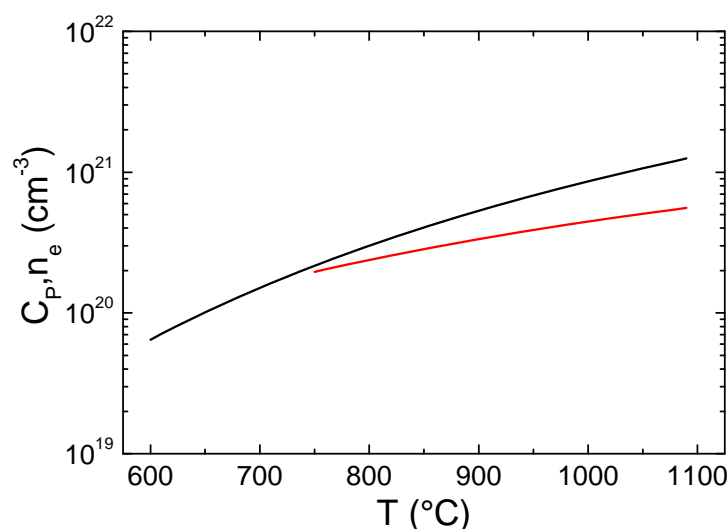


Abbildung 3.1: Temperaturabhängigkeit der Sättigungskonzentration von in Silicium gelöstem Phosphor im Gleichgewicht mit SiP-Präzipitaten (obere Kurve) im Vergleich zur jeweiligen Konzentration der Ladungsträger, die der Konzentration von elektrisch aktivem Phosphor entspricht (untere Kurve) [Solmi '96].

Aus dem Vergleich ergibt sich, dass für Temperaturen oberhalb von 750 °C die Sättigungskonzentration von gelöstem und damit auch mobilem Phosphor die des aktiven übersteigt. Die Konzentration des inaktiven mobilen Phosphors ($C_{sat} - n_e$) steigt auch mit der Temperatur, bei 1000 °C ist ungefähr die Hälfte des gelösten Phosphors inaktiv und trägt somit nicht zur elektrischen Dotierung des Siliciums bei.

3.2 Diffusion von Phosphor in Silicium

Zur Diffusion in Festkörpern finden sich zahlreiche Abhandlungen in der Literatur, wie auch zur Diffusion von Dotierstoffen im Silicium. So sei hier als Beispiel auf eine Einführung in dem Buch von Plummer, Deal und Griffin verwiesen [Plummer '00] und auf einen Review-Artikel von Fahey et al. [Fahey '89]. Im Folgenden soll daher nur auf grundlegende Aspekte eingegangen werden, die für die Beschreibung von Motivation und Ergebnissen der vorliegenden Arbeit von Bedeutung sind.

3.2.1 Beschreibung der Diffusion nach Fick

Eine generelle phänomenologische mathematische Beschreibung der Transportvorgänge von Materie in Festkörpern durch Diffusion ist bei Anwendung grundlegender Gesetze wie Massenerhaltung und daraus abgeleiteter Kontinuitätsgleichung unter Annahme einer statistischen Gesetzmäßigkeiten folgenden Bewegung der Atome im Festkörper möglich, ohne die einzelnen zugrunde liegenden atomaren Vorgänge zu kennen. Bereits Mitte des neunzehnten Jahrhunderts, also lange bevor die atomare Struktur der Festkörper aufgeklärt werden konnte, wurden von Fick die allgemeinen Transportgleichungen, wie sie zur Beschreibung der Wärmeleitung oder Stromleitung bereits bekannt waren, auf die Vorgänge der Diffusion angewandt.

Das erste Fick'sche Gesetz stellt den Zusammenhang von Transport, das heißt Fluss durch Diffusion, und räumlichen Konzentrationsunterschieden dar:

$$J = -D \frac{\partial}{\partial z} C(z, t). \quad (11)$$

Bei Vorhandensein eines Konzentrationsgradienten erfolgt ein Fluss J . Der Proportionalitätsfaktor zwischen Fluss und Konzentrationsgradient ist der Diffusionskoeffizient¹ D . Hierbei ist anzumerken, dass der Fluss im Allgemeinen eine vektorielle Größe darstellt und somit D in diesem Falle ein Tensor ist. In anisotropen Medien oder unter Einfluss äußerer Felder sowie mechanischer Spannungen ist dies zu berücksichtigen. Silicium im spannungsfreien Zustand, wie es hier betrachtet wird, ist jedoch ein isotropes Material von kubischer Struktur, so dass D als Skalar betrachtet werden kann. Ohne weitere Beschränkung der Allgemeinheit wird hier der eindimensionale Fall betrachtet, wie in Gleichung (11) angegeben.

Das Gesetz ergibt sich aus der Annahme einer statistischen gleichverteilten Bewegung der Teilchen. Aufgrund der größeren Anzahl von Teilchen in Bereichen höherer Konzentration erfolgt ein größerer Fluss in Bereiche geringerer Konzentration als umgekehrt. Ein positiver (Netto-)Fluss erfolgt also abwärts des Konzentrationsgradienten, so dass sich das Minuszeichen in der Gleichung ergibt. Bei eingehenderer thermodynamischer Betrachtung zeigt sich

¹ Oft wird D auch als Diffusionskonstante bezeichnet, sie weist jedoch selbst in der Regel eine exponentielle Abhängigkeit von der Temperatur auf.

hier, dass die Ursache für einen Transport letztlich eine thermodynamische Bewegung der Atome ist und der Konzentrationsausgleich die Folge der Maximierung der Entropie des Systems ist.

Das zweite Fick'sche Gesetz beschreibt den Zusammenhang von zeitlicher und örtlicher Veränderung der Konzentration und ist ein Ausdruck des Gesetzes der Erhaltung der Materie in der Kontinuitätsgleichung. Bei der Betrachtung eines Volumenelementes kann sich eine Konzentrationsänderung einer Spezies mit der Zeit darin nur aus der Differenz von hinein- und hinausgehendem Fluss der Spezies ergeben, wenn keine Erzeugung oder Vernichtung der Spezies in dem Element auftritt:

$$\frac{\partial}{\partial t} C(z, t) = - \frac{\partial}{\partial z} J(z, t). \quad (12)$$

Mit Gleichung (11) folgt daraus

$$\frac{\partial}{\partial t} C(z, t) = \frac{\partial}{\partial z} \left(D \frac{\partial}{\partial z} C(z, t) \right), \quad (13)$$

auch bezeichnet als zweites Fick'sches Gesetz (veröffentlicht 1855).

Zwei Lösungen der Differentialgleichung (13) für spezielle (idealisierte) Randbedingungen und ohne äußere und innere Felder und Erzeugungs- und Vernichtungsprozesse der betrachteten Spezies sollen im Folgenden angegeben werden. Bei der Anwendung von technologisch bedeutsamen Diffusionsprozessen können oft unter der Annahme solcher idealen Bedingungen die zu erwartenden Dotierstoffkonzentrationsverteilungen im Festkörper abgeschätzt oder umgekehrt anhand der ermittelten Profilverläufe die beim Prozess vorherrschenden Randbedingungen eingegrenzt werden.

Unter der Annahme eines ortsunabhängigen Diffusionskoeffizienten ergibt sich die vereinfachte Form der Diffusionsgleichung,

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial z^2}. \quad (14)$$

Daraus ergeben sich nun folgende Lösungen.

Eindiffusion aus einer begrenzten Quelle

Dieser Fall liegt vor, wenn eine Diffusion nach einer Oberflächenbelegung mit Dotierstoff z.B. durch Implantation oder Schichtabscheidung aus dieser Quelle heraus erfolgt. Folgende Randbedingungen sollen dabei gelten:

Vor der Diffusion ist noch kein Dotierstoff im Festkörper,

$$C(z, 0) = 0, \quad \forall z > 0. \quad (15)$$

Ein einseitig unendlich ausgedehnter Festkörper wird angenommen,

$$C(\infty, t) = 0, \quad \forall t. \quad (16)$$

Bei den gewöhnlichen Dotierzeiten und den in Silicium vorliegenden Diffusionskoeffizienten ist dies für die Abmessungen der zu dotierenden Materialien in der Regel keine einschränkende Bedingung.

Außer durch Diffusion erfolgt kein weiterer Verlust aus der Quelle, wie z.B. durch Desorption, so dass die in der Quelle an der Oberfläche anfänglich vorhandene Dosis C_D sich nur innerhalb des Festkörpers verteilt,

$$\int_0^{\infty} C(z, t) dz = C_D, \quad \forall t. \quad (17)$$

Die räumliche und zeitliche Entwicklung der Dotierstoffkonzentration im Festkörper ergibt sich dann als gaußförmiger Verlauf,

$$C(z, t) = \frac{C_D}{\sqrt{\pi D t}} e^{-\frac{z^2}{4Dt}}, \quad (18)$$

wobei die Oberflächenkonzentration C_{OF} , d.h. die Konzentration der Quelle, mit der Zeit geringer wird,

$$C(0, t) = C_{OF} = \frac{C_D}{\sqrt{\pi D t}}. \quad (19)$$

Abgeleitet von der gaußförmigen Verteilung wird als wichtige Größe zur Charakterisierung des Transports durch Diffusion die Diffusionslänge definiert,

$$z_D = 2\sqrt{Dt}, \quad (20)$$

als die Tiefe, bei der die Konzentration auf das e^{-1} -fache der Oberflächenkonzentration gefallen ist.

Diffusion aus konstanter Oberflächenquelle (unerschöpfliche Quelle)

Eine Diffusion aus konstanter Oberflächenquelle liegt vor, wenn sich beim Diffusionsprozess ein unveränderliches Gleichgewicht an der Oberfläche des Festkörpers zwischen Dotierstoffaufnahme aus einem praktisch unbegrenzten Vorrat und Dotierstoffabgabe durch Diffusion in den Festkörper hinein einstellt. Bei einer konventionellen Diffusion aus der Gasphase können solche Bedingungen oft vorausgesetzt werden, das sich einstellende Gleichgewicht ist dann bei ausreichender Dotierstoffzufuhr die Festkörperlöslichkeit bei der Diffusionstemperatur.

Die folgenden Randbedingungen können somit formuliert werden:

Die konstante Oberflächenkonzentration an Dotierstoff liegt während der Diffusionszeit vor,

$$C(0, t) = C_{OF}, \quad \forall t. \quad (21)$$

Weiterhin sollen die entsprechenden Bedingungen wie im vorhergehenden Falle gelten,

$$C(\infty, t) = 0, \quad \forall t, \quad (22)$$

$$C(z, 0) = 0, \quad \forall z > 0. \quad (23)$$

Als Lösung ergibt sich nun der folgende Konzentrationsverlauf,

$$C(z, t) = \frac{2C_{OF}}{\pi} \int_{\frac{z}{2\sqrt{Dt}}}^{\infty} e^{-\zeta^2} d\zeta = C_{OF} \operatorname{erfc} \left(\frac{z}{2\sqrt{Dt}} \right), \quad (23)$$

wobei die komplementäre Fehlerfunktion wie folgt definiert ist:

$$\operatorname{erfc}(z) = 1 - \operatorname{erf}(z) = 1 - \frac{2}{\sqrt{\pi}} \int_0^z e^{-\zeta^2} d\zeta. \quad (24)$$

Die Menge des durch Diffusion in den Festkörper eingebrachten Dotierstoffs ergibt sich daraus wie folgt:

$$C_D = \int_0^{\infty} C(z, t) dz = \frac{2C_{OF}}{\sqrt{\pi}} \sqrt{Dt}. \quad (25)$$

Sie wächst mit der Diffusionszeit entsprechend der Diffusionslänge.

3.2.2 Vergleich der Profilform – Abschätzung der benötigten Dosis

Ein Vergleich der beiden Profilformen für verschiedene Diffusionszeiten ist in der folgenden Abbildung angegeben. Für die Berechnung der Gaußverläufe (dicke Linien) der Diffusion aus erschöpflicher Quelle wurde von einer Phosphorquelle an der Oberfläche ausgegangen, die einer Belegung mit einer Monolage entspricht (siehe Kapitel 3.1.3), und für den Diffusionskoeffizienten wurde der intrinsische von Phosphor in Silicium bei 1000 °C gewählt¹:

$$C_D = 6,78 \cdot 10^{14} \text{ cm}^{-2},$$

$$D = 1,32 \cdot 10^{-14} \text{ cm}^2 / \text{s}.$$

Für einen direkten Vergleich wurde für die Berechnung der Verläufe aus unerschöpflicher Quelle jeweils diejenige konstante Oberflächenkonzentration gewählt, die sich für das entsprechende Gauß-Profil ergeben hat.

¹ Siehe dazu folgendes Kapitel 3.2.3. Bei den sich einstellenden Phosphorkonzentrationen wäre hier zwar nicht mehr von einer intrinsischen Diffusion auszugehen, jedoch zur Veranschaulichung der Größenordnungen und vor allem zum Vergleich der Profilformen ist diese Annahme gerechtfertigt.

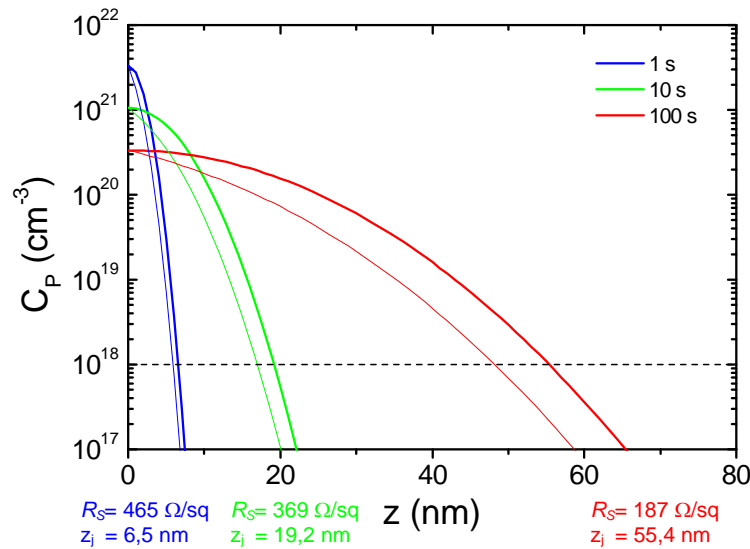


Abbildung 3.2: Vergleich von gaußförmigem Profilverlauf nach Gleichung (18) (dicke Linien) und dem der komplementären Fehlerfunktion nach Gleichung (24) (dünne Linien) für drei verschiedenen Diffusionszeiten für die im Text angegebenen Bedingungen. Dargestellt ist die Phosphorkonzentration über die Tiefe. Für die gaußförmigen Profile sind jeweils die berechneten Flächenwiderstände und die Übergangstiefen bei $C_P = 1 \cdot 10^{18} \text{ cm}^{-3}$ angegeben.

Der Vergleich der Profile zeigt, dass bei jeweils gleicher Oberflächenkonzentration der gaußförmige Verlauf einem ideal rechteckförmigen näher ist, was günstiger ist für das Erreichen der geforderten pn-Eigenschaften, wie in Kapitel 2.2.2.3 beschrieben. Die Abbildung veranschaulicht außerdem die Größenordnung der Dotierung, die sich mit einer Belegung von nur einer Monolage Phosphor auf der Oberfläche (idealerweise) erzielen ließe, wenn kein Verlust aus dieser Quelle durch Phosphordesorption oder Trapping an der Grenzfläche aufträte. Unter der Annahme einer vollständigen Aktivierung der Monalage nach Eindiffusion zu gaußförmigem Profilverlauf wurden mittels der Simulationssoftware ATHENA und ATLAS (Fa. SILVACO) die Flächenwiderstände berechnet. Sie sind gemeinsam mit den Übergangstiefen bei einer Konzentration von $C_P = 1 \cdot 10^{18} \text{ cm}^{-3}$ in der Abbildung 3.2 für das jeweilige Gaußprofil angegeben. Aus diesen Werten lässt sich ablesen, dass selbst die hohen Anforderungen an die Dotierung der SDE-Bereiche (siehe Abbildung 2.4) mit einer Belegung von nur einer Monolage bei geeigneter Eindiffusion und Aktivierung prinzipiell erfüllbar sein sollten.

Die Menge des dafür benötigten Phosphors lässt sich einfach abschätzen. Bei der Verwendung von Wafern mit 150 mm Durchmesser ist die Waferoberfläche ca. 175 cm^2 groß. Bei einer benötigten Dosis für eine Monolage von $6,78 \cdot 10^{14} \text{ cm}^{-2}$ ergibt dies einen Bedarf von ca. $1,2 \cdot 10^{17}$ Phosphoratomen.

3.2.3 Diffusionskoeffizienten von Phosphor in Silicium

Die Diffusionskoeffizienten für diese Modelle zeigen eine exponentielle Abhängigkeit von der Temperatur in Form eines Arrheniusgesetzes,

$$D = D_0 e^{-\frac{E_A}{kT}}, \quad (26)$$

mit einer Aktivierungsenergie E_A für die Diffusion, wobei der exponentielle Term die thermodynamische Wahrscheinlichkeit angibt, dass das diffundierende Teilchen eine Energie oberhalb von E_A besitzt. Diese Aktivierungsenergie liegt für Dotierstoffe in Silicium im Be-

reich zwischen 3,5 und 4,5 eV. Im einfachen Potentialmodell des Festkörpers entspricht diese Aktivierungsenergie der Potentialbarriere zwischen zwei gebundenen Zuständen des Dotierstoffatoms mit dem Festkörper, der für einen räumlichen Übergang überwunden werden muss. Die Größe D_0 in (27) wird als Konstante betrachtet, obwohl sie selbst eigentlich eine gewisse Temperaturabhängigkeit aufweist, sie ist ein Maß für die Frequenz (Debye-Frequenz bzw. Gitterschwingung), mit der die Atome die Potentialbarriere zu überwinden vermögen.

Diese einfache Temperaturabhängigkeit der Diffusionskoeffizienten gilt für intrinsisches, eigenleitendes Silicium bei der entsprechenden Diffusionstemperatur, wenn eine Abhängigkeit durch die internen elektrischen Felder diffundierender geladener Dopanten zu vernachlässigen ist. Für eine n-Dotierung in p-Material muss für die Konzentration N_D von einzubringenden Donatoren also gelten:

$$n_i > \frac{N_D}{N_A}. \quad (27)$$

Die intrinsische Ladungsträgerkonzentration n_i in Silicium kann nach folgender Formel berechnet werden [Morin '54],

$$n_i = 3,9 \cdot 10^{16} \cdot e^{-\frac{0,605 \text{ eV}}{kT}} \cdot T^{1,5} \text{ cm}^{-3}, \quad (28)$$

und ist für die gewöhnlichen Diffusionstemperaturen bereits im Bereich von mittleren Dotierstoffkonzentrationen, wie aus der folgenden Abbildung ersichtlich wird.

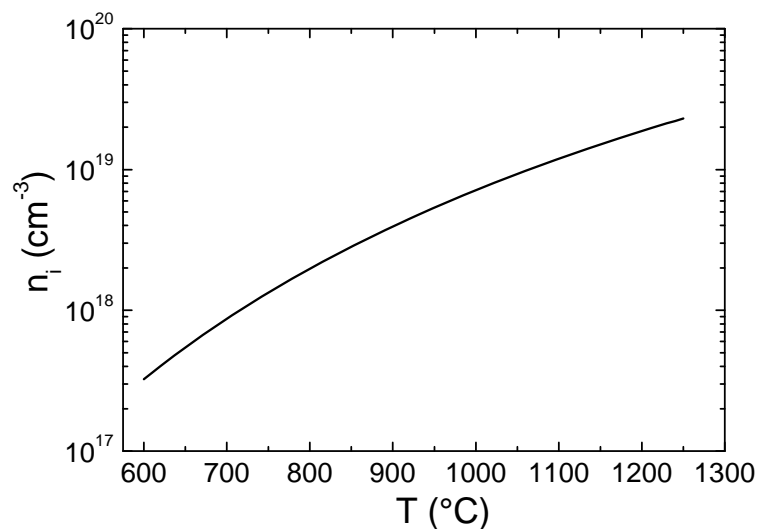


Abbildung 3.3: Temperaturabhängigkeit der intrinsischen Ladungsträgerkonzentration in Silicium nach Gleichung (28) aus [Morin '54].

Eine exakte Bestimmung des intrinsischen Diffusionskoeffizienten erweist sich als schwierig und oft abhängig von den jeweiligen experimentellen Bedingungen, aufgrund der sehr vereinfachenden Annahmen der Modellbildung. Für die Phosphordiffusion in Silicium werden z.B. die Tabelle 3.1 angegebenen Werte für die Beschreibung der intrinsischen Diffusion verwendet.

Für höhere Konzentrationen von Phosphor als die intrinsische zeigt sich, dass der Diffusionskoeffizient auch konzentrationsabhängig ist (extrinsischer Fall). Eine Beschreibung kann dann in der Form

$$D_p = D^0 + D^- \left(\frac{n}{n_i} \right) + D^= \left(\frac{n}{n_i} \right)^2 \quad (27)$$

erfolgen, wobei D^0 , D^- und $D^=$ als die Diffusionskoeffizienten des Phosphors, jeweils gepaart mit neutralen, einfach negativen und zweifach negativen Punktdefekten aufgefasst werden. Für die Phosphordiffusion wird hierbei in der Regel von einer Paarung mit Interstitials ausgegangen. So zeigen Diffusionsexperimente unter gleichzeitiger Oxidation, die mit einer Injektion von Interstitials ins Silicium verbunden ist, eine Diffusionsverstärkung (OED - oxidation enhanced diffusion), während bei gleichzeitiger Nitridierung eine Verzögerung der Diffusion beobachtet wurde [Fahey '89].

Tabelle 3.1: Parameter des intrinsischen Diffusionskoeffizienten von Phosphor in Silicium.

D_0 (cm ² /s)	E_A (eV)	Quelle
1,3	3,535	[SILVACO '05]
0,0008	2,74	[Christensen '03]
0,00171	2,81	[Haddara '00]
4,7	3,68	[Plummer '00]
5,30	2,81	[Uematsu '97] aus [Makris '73]
0,074	3,3	[Ghoshtagore '71]

Bei der Beschreibung über Gleichung (27) folgt für den intrinsischen Fall,

$$n = n_i, \quad (28)$$

für den effektiven Diffusionskoeffizienten des Phosphors¹,

$$D_p = D^0 + D^- + D^=. \quad (29)$$

Die einzelnen Diffusionskoeffizienten sind wiederum entsprechend dem Arrheniusgesetz (26) von der Temperatur abhängig. Literaturwerte dieser Koeffizienten sind in der Tabelle 3.2 angegeben.

¹ Von der effektiven Diffusion wird außerdem gesprochen, wenn die Temperatur während der Diffusion nicht konstant ist oder wenn mehrere Hochtemperaturschritte aufeinander folgen und diese somit das letztendliche, effektive Diffusionsprofil beeinflusst haben, was bei praktischen Anwendungen in der Regel der Fall ist. Das effektive Dt -Produkt ist ein Maß für das *thermische Budget*, es ist die Summe aller einzelnen Dt -Produkte:

$$(Dt)_{eff} = \sum D_i t_i. \quad (30)$$

Da D exponentiell von der Temperatur abhängig ist, dominiert der Einfluss des Temperaturschritts mit der höchsten Temperatur (auch bei verhältnismäßig kürzerer Zeit) auf das entgültige Konzentrationsprofil.

Tabelle 3.2: Parameter der extrinsischen Diffusionskoeffizienten von Phosphor in Silicium.

D^0		D^-		$D^=$		Quelle
D_0^0 (cm ² /s)	E_A^0 (eV)	D_0^- (cm ² /s)	E_A^- (eV)	$D_0^=$ (cm ² /s)	$E_A^=$ (eV)	
3,85	3,66	4,44	4,00	44,2	4,37	[SILVACO '05]
0,00021	2,65	0,47	3,5			[Wittel '95]

Eine vergleichende Darstellung der Temperaturabhängigkeit der in den Tabellen angegebenen Diffusionskoeffizienten ist in Abbildung 3.4 ersichtlich.

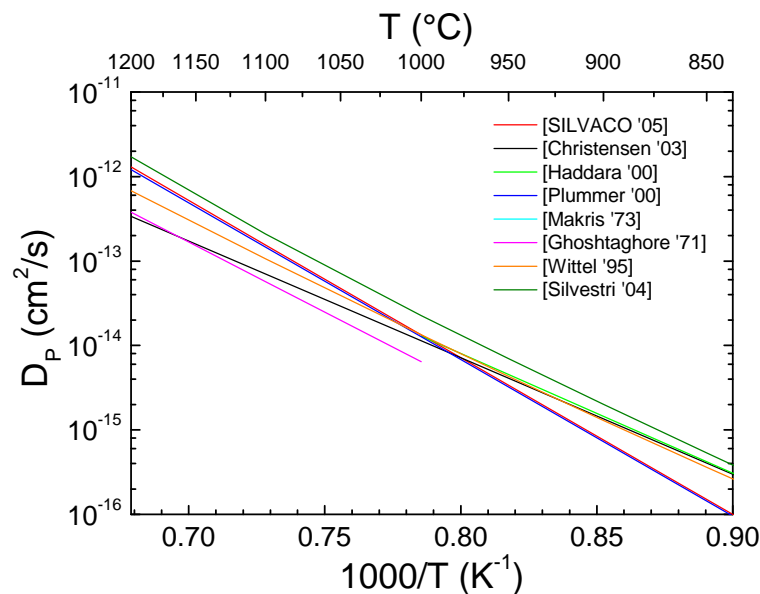


Abbildung 3.4: Vergleich von Literaturwerten des intrinsischen Diffusionskoeffizienten von Phosphor in Silicium anhand der logarithmischen Darstellung über der reziproken Temperatur (Arrhenius-Darstellung).

Der Vergleich von Literaturwerten zeigt auch bei den extrinsischen Diffusionskoeffizienten, dass eine exakte Bestimmung schwierig ist und Unterschiede in Anhängigkeit von den Untersuchungsbedingungen auftreten, z.B. aufgrund von Paarungsreaktionen des Phosphors mit einer Bor-Grunddotierung [Wittel '95]. Auch der genaue Mechanismus der Paardiffusion ist noch nicht vollständig aufgeklärt. So ergibt sich z.B. bei den Untersuchungen von Wittel und Dunham kein Hinweis auf eine Diffusion über doppelt negativ geladene Defekte. Während neuere detaillierte Analysen von Paarungsreaktionen im Zusammenhang mit den Untersuchungen der Eigendiffusion in Silicium auch die Diffusion von positiven Paaren von Phosphor mit Interstitials ohne das Vorhandensein eines Leerstellenmechanismus postulieren [Silvestri '04], werden in anderen Modellen auch die Paarungen mit Leerstellen als erforderlich für eine vollständige Beschreibung der Phosphordiffusion angenommen [Yoshida '02]. Dies schließt einige besondere Phänomene, die bei der Phosphordiffusion bei sehr hohen Phosphorkonzentrationen oder in Anwesenheit einer Übersättigung mit Defekten, insbesondere nach einer Vorbelegung durch Implantation, gefunden wurden, ein, so z.B. die „phosphorus tail diffusion“, das „Abknicken“ des Dotierprofils aufgrund einer verstärkten Diffusion im Bereich geringerer Phosphorkonzentrationen durch transiente Transportphänomene von Defekten in diesen Bereich.

Auf weitergehende Ausführungen zu den Diffusionsvorgängen soll hier jedoch verzichtet werden, da die obigen Modelle zur Beschreibung der untersuchten Diffusionsprozesse ausreichen werden. Für ausführlichere Darstellungen sei auf die zitierte Literatur verwiesen.

4. Experimentelle Vorgehensweise

4.1 RTP-Modul

Für die experimentellen Untersuchungen wurde ein RTP-Cluster-Modul Hector 300 von AST (später STEAG, heute Firma MATTSON) verwendet. Dabei handelte es sich ursprünglich um ein RTCVD-Modul (rapid thermal chemical vapor deposition), das als Prototypenanlage für Entwicklungsaufgaben im Einsatz war und mehrfach modifiziert wurde. Im Folgenden sollen kurz wichtige Parameter des Moduls vorgestellt werden.

Prozesskammer

Das Modul besitzt eine Prozesskammer aus Edelstahl, in der sich ein Quarzgestell für die Aufnahme eines Wafers von 150 mm Durchmesser befindet und die mittels Vakuumpumpen evakuiert werden kann (siehe Abbildung 4.1). Über ein Rechteckventil ist die Prozesskammer mit einer zentralen Cluster-Transferkammer verbunden. Die Prozessgaszuführung befindet sich oberhalb der Beladeöffnung, zwei separate Einlässe sind als kammförmiges Düsensystem ausgebildet. An der gegenüberliegenden Seite befindet sich der Absaugflansch. Mit dieser Anordnung wird ein möglichst homogenes Anströmen der Waferoberfläche mit Prozessgas angestrebt. An der Kammerunterseite befindet sich eine Quarzglasplatte, durch die die Strahlung der Lampen in die Kammer gelangen kann. Die Kammerwände sind wassergekühlt, die Kammerinnenwände sind nicht poliert oder verspiegelt.

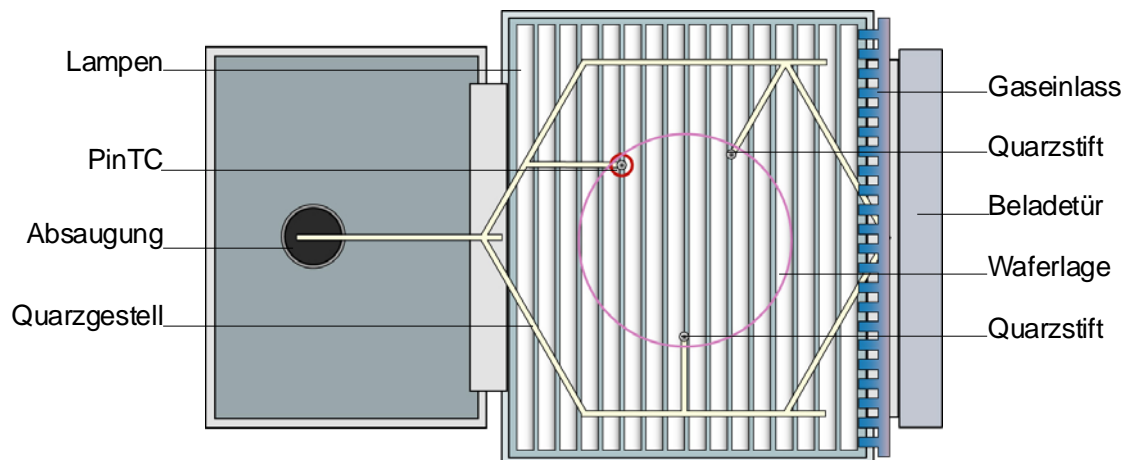


Abbildung 4.1: Schematische Darstellung der Anordnung von Komponenten des RTP-Moduls (von oben betrachtet)

Waferhalter

Um eine geringe thermische Trägheit der Waferaufnahme und möglichst ganzflächige Bestrahlung der Waferunterseite zu gewährleisten, liegt der Wafer in der Kammer auf drei Stiften (pins) auf, die sich auf einem Quarzgestell (tray) befinden. Zwei von diesen Stiften sind aus Quarzglas, der dritte ist als so genannter PinTC ausgeführt, so dass diese Auflage eine Temperaturmessstelle in direktem Kontakt mit dem Wafer darstellt [Lerch '98].

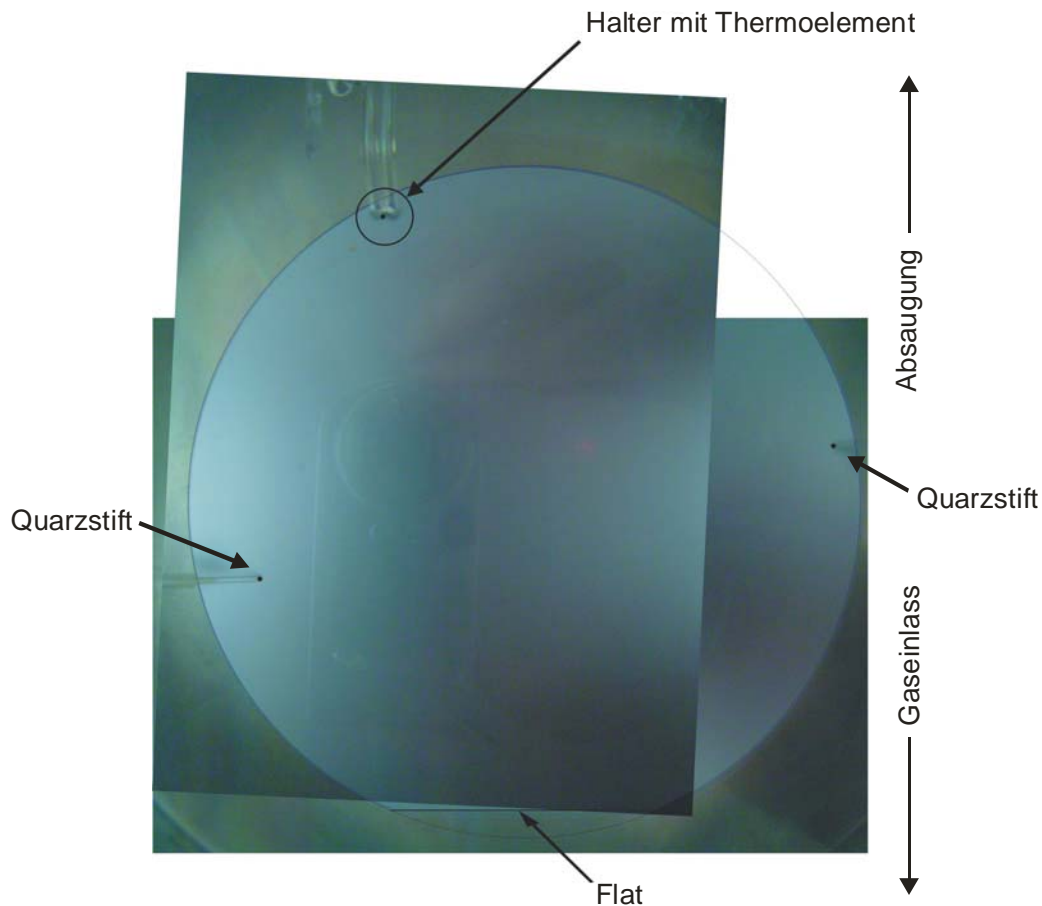


Abbildung 4.2: Lage eines Wafers auf drei Auflagepunkten, einer davon mit Thermoelement (Eigenbau), Ansicht von unten durch die Quarzplatte (aus zwei Fotos zusammengesetzt).

Temperaturmessung und -regelung

Die Wafertemperatur wird mittels eines Thermoelements (TC - thermo couple) vom Typ K im PinTC gemessen. Das PinTC gestattet eine direkte Temperaturmessung am Wafer, ohne dabei die optischen Emissionseigenschaften des Wafers kennen zu müssen. Dies ist besonders bei Waferoberflächen von Vorteil, die durch Beschichtungen oder Strukturierungen, welche auch beim zu kontrollierenden Prozess selbst entstehen oder beeinflusst werden, unbestimmte oder sich ändernde optische Eigenschaften besitzen. Damit ist eine optische, pyrometrische Temperaturerfassung problematisch. Für die hier untersuchten Dotierprozesse sollte dies zwar keine Beschränkung darstellen, ein Pyrometer für eine optische Temperaturerfassung konnte aufgrund der speziellen Ausführung der Prozesskammer jedoch nicht genutzt werden.

Allerdings erwies sich das PinTC während der Versuchsreihen als sehr fragil und wenig belastbar. Es kam wiederholt zu Ausfällen des PinTCs, wobei Reparaturen nicht möglich waren und ein Austausch erforderlich wurde. Aufgrund der hohen Kosten, die mit der Anschaffung des Original-PinTC verbunden waren, wurde ein Ersatz entwickelt, bei dem bei Bedarf unter weiterer Verwendung des Halters das Thermoelement ausgetauscht werden konnte. Dafür wurden unisolierte Thermoelementdrähte in zwei verbundenen Quarzkapillaren geführt. Die Messstelle an der Auflagespitze wurde mit Keramikzement aus Aluminiumoxid (Resbond 903HP von Fa. COTRONICS) fixiert und elektrisch isoliert. Auf eine Siliciumcarbidkappe, mit der das Original-PinTC ausgestattet ist, wurde verzichtet.

Die Regelung der Temperatur erfolgte durch ein schnelles geschlossenes Regelsystem (closed-loop), für das die Messgröße des PinTCs den Echtzeit-Temperatur-Istwert lieferte. Abweichungen dieses Istwertes vom vorgegebenen Sollwert wurden über die Veränderung der Regelgröße Lampenleistung minimiert. Die Regelparameter (Proportionalwert-P, Integralwert-I und Differentialwert-D) konnten für jeden Prozessschritt separat festgelegt werden. Damit konnten für die verschiedenen Bereiche des Temperaturprofils die Anpassung von Ist- und Sollwertverlauf optimiert werden. Diese Regelung erwies sich als zuverlässig und stabil, so dass für alle Prozesse PID-Parameter wie in Tabelle 4.2 angegeben verwendet werden konnten.

Lampen

Der Siliciumwafer wird durch Strahlung aus einem Feld von 16 Wolfram-Halogen-Stablampen mit je 1500 Watt Nennleistung beheizt. An der Anlage ist nur dieses eine Lampenfeld vorhanden, das unterhalb des Wafers außerhalb des Reaktors angeordnet ist. Der Boden der Reaktorkammer besteht aus einer Quarzglasplatte, durch die die Lampenstrahlung in das Innere der Kammer gelangt. Die Lampen sind in drei Gruppen (lamp banks), im Verhältnis 5:6:5 verschaltet, siehe Abbildung 4.3.

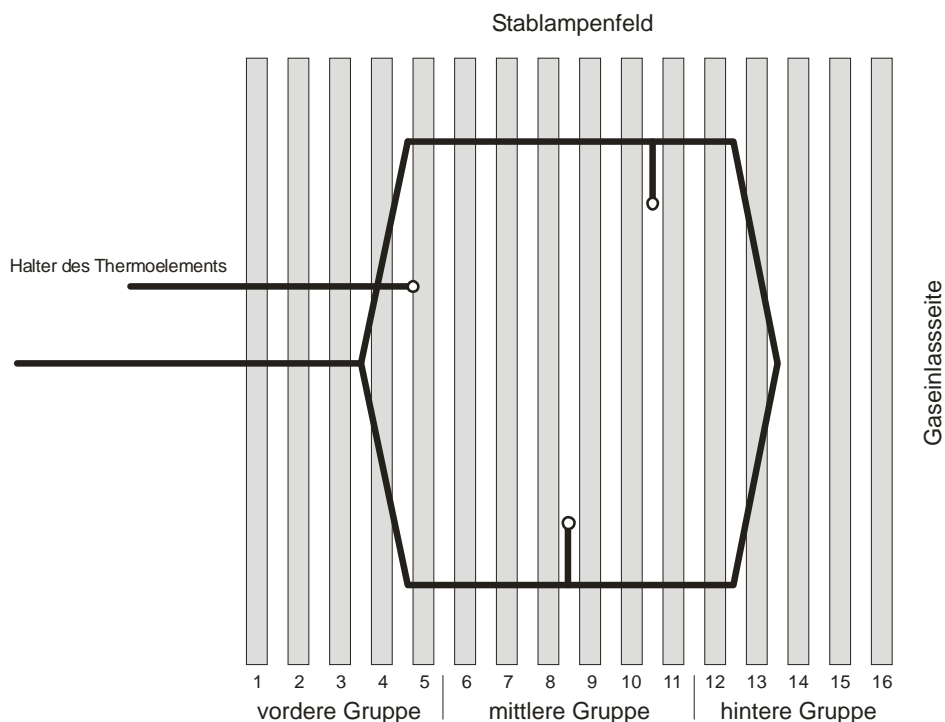


Abbildung 4.3: Schematische Darstellung der Anordnung der Lampen und des Quarzgestells der Waferauflage (Ansicht von oben).

Eine Gewichtung der Verteilung Leistung der Lampen kann über die Rezeptsteuerung nur für die Gruppen separat erfolgen. Eine Leistungswichtung jeder einzelnen Lampe ist über Korrekturtabellen möglich (lamp correction table), in denen für jede Lampe ein Gewichtungsfaktor abgelegt ist. Diese Tabellen können nicht über die Rezeptbedienung erstellt oder verändert werden, jedoch kann im Rezept die Verwendung vorhandener Korrekturtabellen für jeden einzelnen Prozessschritt festgelegt werden. Da die Anlage jedoch nicht über obere oder seitliche Lampenfelder verfügt, ist eine Optimierung der Temperaturverteilung auf dem Wafer nur in sehr eingeschränktem Maße möglich.

Kühlung

Die Edelstahlwände der Reaktorkammer sind wassergekühlt. Zusätzlich sind Teile der Leistungselektronik der Lampenansteuerung und die Lampenfeldplatte aus Aluminium, auf der die Lampen montiert sind und deren Oberfläche als polierte Reflektorenanordnung ausgebildet ist, in einen weiteren Kühlkreislauf eingebunden. Die Quarzglasplatte wird durch einen Luftstrom gekühlt, der in einem Gebläse erzeugt wird und durch Öffnungen der gekühlten Lampenfeldplatte geführt wird. Die Gebläseleistung kann über die Rezeptsteuerung angepasst werden, eine Temperaturmessstelle an der Quarzplatte ist bei der verwendeten Anlage jedoch nicht vorhanden. An der Kammeraußenseite oberhalb der Lampen wurden Wandtemperaturen bis zu ca. 70 °C gemessen.

Gasversorgungs- und Vakuumsystem

Das RTP-Modul ist für das Ausführen von Prozessen im Vakuum ausgelegt. Die Vakuumerzeugung erfolgt mittels einer einstufigen Trockenläuferpumpe vom Typ ADP 81 (Fa. ALCA-TEL). Für Druckmessung und -regelung sind Kapazitätsmanometer und Butterflyventil vorhanden. Mit dieser Ausstattung lässt sich ein Enddruck von 0,02 mbar erreichen und bei maximalem Gaseinlass von 100 sccm ein stationärer Druck von bis zu ca. 5 mbar im Prozessraum präzise einstellen. Zusätzlich wurde für einige Versuche ein mobiler Turbomolekularpumpstand an den Reaktor angeschlossen, der nicht in die Anlagensteuerung eingebunden war und manuell bedient wurde. Dieser wurde zur Konditionierung der Kammer vor Prozessbeginn eingesetzt; damit wurden Enddrücke, gemessen mit Penning-Vakuummeter, bis hinunter zu ca. $1 \cdot 10^{-5}$ mbar erreicht.

Die Gaszuführung erfolgt aus der zentralen Infrastruktur des Reinraums. Es wurden nur hochreine Gase in Halbleiterqualität verwendet. Die Zuführung erfolgte in elektropolierten und ausschließlich verschweißten oder metallgedichteten Leitungen. Für die Experimente wurden in folgender Tabelle aufgeführte Gaslinien verwendet.

Tabelle 4.1: **Verwendete Gaslinien und der jeweils maximal mittels Massendurchflussmesser zu regelnde Gasfluss.**

Gasart	Maximalfluss
Stickstoff	15 slm
Sauerstoff	100 sccm
Wasserstoff	50 sccm
Phosphingemisch	100 sccm

Die Sauerstofflinie ist bis hin zum Einlass in die Prozesskammer von den übrigen getrennt. Um reproduzierbare und stabile Gasflüsse in die Prozesskammer schalten zu können, ist ein Bypass-System vorhanden, das eine Gasflusseinstellung direkt in den Abgasstrang zur Pumpe

erlaubt, bei Prozessgasbedarf kann dieser Gasfluss durch die Kammer geleitet werden. Sämtliche Gaseinlässe und -mengen während der Prozesse werden mit der Rezeptbearbeitung automatisch gesteuert. Von der Vakuumpumpe wird das Abgas einem Abgasverbrenner und -wäscher Flawamat (Fa. CENTRO THERM) zugeführt.

Prozesssteuerung und Datenaufzeichnung

Die Steuerung der Prozesse erfolgt ausschließlich über die schrittweise Abarbeitung von vorher erstellten Rezepten. Hierin lassen sich alle für den Prozessablauf relevanten Steuerbefehle ablegen. Die Prozesszeit wird durch die festgelegte Länge der Einzelschritte bestimmt oder es können Sprung- oder Abbruchbedingungen definiert werden. In der folgenden Tabelle ist ein Beispiel für die im Rezept einzustellenden Parameter eines typischen Annealing-Prozesses aufgeführt.

Tabelle 4.2: Rezept in tabellarischer Form für ein 10 s - Annealing in 5 mbar Sauerstoff bei 900 °C Solltemperatur.

RECIPE NAME : 1_AO_900.011
 AUTHOR : Bodo
 FACILITY : Uni MD
 CALIBRATION : emis1300

STEP	STEP	LINE	TEMP	TEMP	TEMP	TEMP	TEMP	BANK	BANK	BANK	LINE	LINE	LINE	LINE	LINE	LINE	LINE	DIGOUT	DIGOUT	C.HOLD	C.HOLD	C.HOLD	
step name	time	Temperature	P	I	D	RESET	LCT	FrontBan	CentreBn	BackBank	Blower1	N2	O2	NO	H2	B2H6	PH3	Pressure	GasHydride	GasHalide	Param.	Value	time
no_step	5	0	0	0	0	0	corr1	0	0	0	0	0	0	0	0	0	0	0,01	OFF	OFF	none	0	0
init_O2	20	0	0	0	0	0	corr1	0	0	0	0	0	100	0	0	0	0	0,01	OFF	OFF	none	0	0
set_O2-press	5	0	0	0	0	0	corr1	0	0	0	0	0	100	0	0	0	0	0,01	OFF	ON	none	0	0
set_temp	60	400	0	0	0	0	sinopt3	20 R	20 R	20 R	80	0	100	0	0	0	0	90	OFF	OFF	TmpRdo	400	0,1
ramp_temp	15	550 R	5	1	0	0	sinopt3	90	90	90	80	0	100	0	0	0	0	90	OFF	ON	none	0	0
prebake	60	550	5	3	0	0	sinopt3	90	90	90	80	0	100	0	0	0	0	90	OFF	ON	none	0	0
ramp	15	900 R	5	3	0	0	sinopt3	90	90	90	80	0	100	0	0	0	0	90	OFF	ON	none	0	0
doping	10	900	5	3	0	0	sinopt3	90	90	90	80	0	100	0	0	0	0	90	OFF	ON	none	0	0
ramp_down	30	400 R	5	1	0	0	sinopt3	90	90	90	80	0	100	0	0	0	0	90	OFF	ON	none	0	0
cooling	180	0	0	0	0	0	corr1	0	0	0	80	0	0	0	0	0	0	90	OFF	OFF	none	0	0
purge_n2	60	0	0	0	0	0	corr1	0	0	0	0	3	0	0	0	0	0	20	ON	OFF	none	0	0
pump_down	30	0	0	0	0	0	corr1	0	0	0	0	0	0	0	0	0	0	0,01	OFF	OFF	none	0	0
end	5	0	0	0	0	0	corr1	0	0	0	0	0	0	0	0	0	0	0,01	OFF	OFF	none	0	0

Rezeptschrittnamen	Schrittdauer in s	Soll-Temperatur in °C (fr: Rampe)	P-, I-, D-Regelparameter für Lampenleistung	zu verwendende Lampenkennwertabelle	Leistungssteigerung auf die Lampengruppen in % (fr: Rampe)	Gebbläseleistung in %	Gasflüsse in sccm oder slm	Prozessdruck in mbar	Schaltung Gaslinien (ohne Sauerstoff) auf die Prozesskammer	Schaltung Gaslinien (mit Sauerstoff) auf die Prozesskammer	Parameter für bedingte Sprünge
--------------------	-------------------	-----------------------------------	---	-------------------------------------	--	-----------------------	----------------------------	----------------------	---	--	--------------------------------

Der Prozessablauf wird automatisch protokolliert und abgespeichert. In einer Initialisierungsdatei können die aufzuzeichnenden Parameter und die Sampling-Rate festgelegt werden. Die wichtigsten Parameter sind in der Tabelle 4.3 aufgelistet.

Tabelle 4.3: Übersicht über die protokollierten Prozessparameter.

Parameter	Einheit	Sampling-Rate
Temperatur-Sollwert	°C	0,1 s
Temperatur-Istwert am PinTC	°C	0,1 s
Druck	mbar	1 s
Gesamtleistung	kW	1 s
Gasflüsse	sccm oder slm	1 s
Leistung der Lampengruppen	0,01 kW	0,1 s

Ein Beispiel für den in Tabelle 4.2 definierten Prozessablauf anhand wichtiger protokollierter Daten ist in der Abbildung 4.4 angegeben.

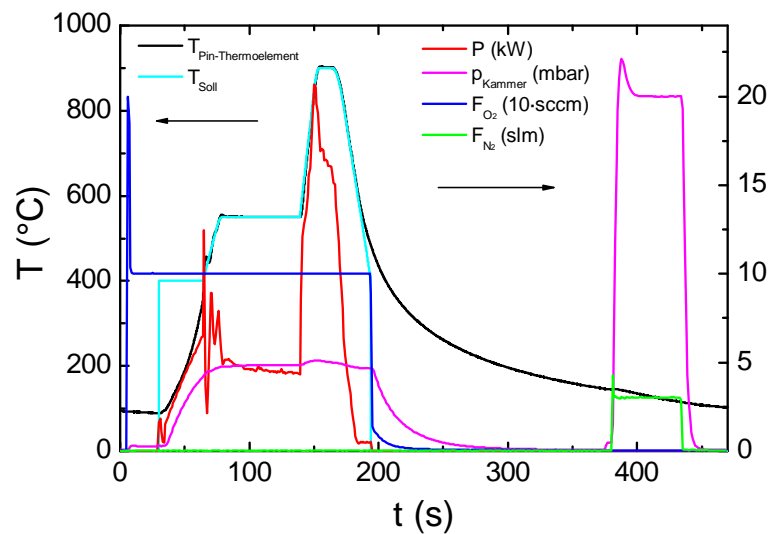


Abbildung 4.4: Darstellung einiger relevanter Prozessvariablen (Vorgabe der Temperatur, T_{Soll} , mit Pin-Thermoelement gemessene Temperatur, $T_{\text{Pin-Thermoelement}}$, sowie Gesamtleistung der Lampen, P , Kammerdruck, p_{Kammer} und Gasflüsse von O_2 , F_{O_2} , und N_2 , F_{N_2}) über der Prozesszeit t für den in der Tabelle 4.2 festgelegten Prozess.

Scheibentransfer

Das RTP-Modul ist Teil eines Anlagenclusters, bestehend aus einer Schleusenkammer (load-lock) und einer zentralen Verteilerstation (dealer oder handler) des Typs AVC 64 (Fa. ASYS GmbH) für das Waferhandling. In diesem Cluster befinden sich weiterhin ein Schichtabscheidungs- und ein Reinigungsmodul. Alle Module des Clusters sind durch Rechteckventile vakuumdicht mit dem Handler verbunden.

Die Schleusenkammer wird über den Reinraum mit einer Waferkassette bestückt, in der sich bis zu 25 Wafer befinden können. Der Handler besitzt für den Wafertransport einen programmierbaren Roboterarm mit sehr genauen Schrittmotoren. Die Bedienung erfolgt über eine Software, die auf einem PC installiert ist. Die Schleusen- und die Handlerkammer werden mittels einer Scroll-Vakuumpumpe evakuiert, das Fluten erfolgt mit Stickstoff. Für die Dotierexperimente erfolgte der Wafertransport im Vakuum bei einem Druck von ca. 0,02 mbar.

4.2 Eingesetzte Materialien

4.2.1 Verwendete Gasgemische

Für die Experimente standen zwei verschiedene Gasgemische zur Verfügung, die Phosphin in unterschiedlicher Konzentration enthielten:

- 9,91 Volumenprozent Phosphin in Stickstoff und
- 52,3 vpm (Volumenanteile pro Million) Phosphin in Wasserstoff.

In den folgenden Ausführungen werden der Einfachheit halber diese Mischungen mit 10 % PH_3 in N_2 bzw. 50 vpm PH_3 in H_2 bezeichnet.

4.2.2 Substratauswahl und -vorbehandlung

Substrate

Für die Dotierversuche wurden ausschließlich (100)-orientierte p-Typ-Siliciumwafer mit einer Bor-Grunddotierung und von 150 mm Durchmesser verwendet, die einen spezifischen elektrischen Widerstand ρ von 8-12 Ωcm aufwiesen. Die p-Typ-Wafer wurden bevorzugt, um im Substrat durch die Dotierung einen pn-Übergang zu erzeugen. Für einzelne Versuchsreihen wurden mehrere Scheiben zu einem Los zusammengestellt.

Reinigung

Für die Standard-Eingangsreinigung erfolgte zunächst ein Bad in Caro'scher Säure (Peroxo-monoschwefelsäure). Sie wurde durch die Mischung von 2 Teilen konzentrierter Schwefelsäure (96%) und einem Teil Wasserstoffperoxid (30%) im Reinigungsbecken dargestellt:



Sie wirkt stark oxidierend auf organische Kontaminationen, die dadurch von der Siliciumoberfläche gelöst werden, und die Oberfläche selbst wird über die Bildung von Silanolgruppen (Si-OH) chemisch oxidiert. Die Standardreinigung erfolgte für 5 min in 120 °C heißer Caro'scher Säure.

Anschließend wurden die Scheiben in DI-Wasser bis auf einen Leitwert des Wassers von $6,25 \cdot 10^{-8}$ S, mindestens jedoch 30 Minuten, gespült.

Zur Entfernung des chemischen Oxids erfolgte ein Eintauchen der Scheiben für eine Minute in einprozentiger wässriger Flusssäurelösung bei Raumtemperatur:



Zur Entfernung von natürlichen Oxiden wurde solch ein HF-Dip auch entsprechend angewendet. Mit der Oxidschicht wurden über die Bildung von löslichen Metallfluoriden auch die metallischen Kontaminationen entfernt, die in der Oxidschicht gebunden waren.

Dann wurde wieder ein Spülschritt mit DI-Wasser auf Leitwert, mindestens aber 30 min, durchgeführt. Das Trockenschleudern der Scheiben erfolgte mit maximal 2000 U/min unter Stickstoffatmosphäre und bei erhöhter Temperatur. Für anschließende Dotierversuche wurden die Scheiben unmittelbar nach der Reinigung in einer Horde in die Schleusenkammer des Clusters geladen und die Kammer wurde evakuiert.

Für Voruntersuchungen zur Reinigung wurde auch eine zum Caro-Bad alternative Vorbehandlung in RCA-SC1 (standard clean 1 der Radio Corporation of America) zur Entfernung von organischen Verunreinigungen eingesetzt. Dieses auch APM (ammonia peroxide mixture) genannte Bad wurde durch Mischung von Ammoniumhydroxid, Wasserstoffperoxid und DI-Wasser im Verhältnis 1:1:5 gewonnen.

4.3 Prozessführung - Auswahl der Experimente

4.3.1 Untersuchungen zur Reinigung und Vorbehandlung der Wafer

Da der angestrebte Dotierprozess über einen Oberflächen-Anlagerungsprozess abläuft, ist eine möglichst präzise Kenntnis der Oberflächeneigenschaften, insbesondere von Reinheit und eventueller Oxidbedeckung wünschenswert. Außerdem sind für die Untersuchungen zur Temperaturverteilung auf den Wafern über eine Kurzzeitoxidation möglichst genaue Informationen zur anfänglichen Bedeckung der Oberfläche und zur Reproduzierbarkeit der Oberflächeneigenschaften der eingesetzten Testwafer nötig.

Daher wurden spezielle Untersuchungen zu Eingangsreinigungen und zur Entfernung von Oxidschichten durchgeführt. Diese Oxidschichten entstehen durch die oxidierenden Reinigungsverfahren, über natürliche Oxidation an der Reinraumluft und auch während der Spülung nach dem jeweiligen Reinigungsschritt. Ziel dieser Untersuchungen war, diese Einflüsse auf die Siliciumoberfläche, wie sie auch unter den speziellen gegebenen Laborbedingungen vor den Dotierprozessen vorlagen, zu charakterisieren.

Dafür wurde ein in der Tabelle 4.4 dargestellter Ablauf der Experimente gewählt.

Tabelle 4.4: Ablauf der Untersuchungen zur nasschemischen Reinigung und Terminierung der Waferoberflächen und zum Wachstumsverhalten natürlicher Oxidschichten.

Schritt	Bemerkungen
Reinigung	Verwendung von verschiedenen Reinigungslösungen
Spülen in DI-Wasser, Schleudern in N ₂	Entfernen von Rückständen der Reinigungslösung und der noch vorhandenen Verunreinigungen, Trocknen
Schichtdickenmessung	Messung am Ellipsometer
Oxidentfernung	HF-Dip verschiedener Konzentrationen und Dauer
Spülen in DI-Wasser, Schleudern in N ₂	Entfernen von an der Oberfläche vorhandenen Fluorverbindungen, Trocknen
Schichtdickenmessung	Zeitaufgelöst am Ellipsometer, Untersuchung des Aufwachsens vom natürlichem Oxid
Wiederholungen der Oxidentfernung nach natürlicher Oxidation und anschließende Schichtdickenmessungen zur Untersuchung der Reproduzierbarkeit der Entfernung und des Wachstums	

Es wurden jeweils unterschiedliche Reinigungsschritte und Schritte zur Oxidentfernung angewendet. Als Reinigungsschritte wurden die Standard-Caro Reinigung und die Reinigung in SC1 durchgeführt. Die Oxidentfernung erfolgte immer mittels HF-Dip, wobei auf deren Reproduzierbarkeit und den Einfluss der Dauer der HF-Behandlung geachtet wurde

Die Reinigungs- bzw. Ätzverfahren sollten dahingehend verglichen werden, ob Unterschiede in Oxidbedeckung und erneutem Oxidwachstum nach der Entfernung zu beobachten waren. Im Hinblick auf diese Beobachtungsziele wurden gegenübergestellt:

- Vergleich der unterschiedlichen Reinigungslösungen,
- Wiederholte Ausführung der HF-Behandlung nach Wachstum von natürlichem Oxid,
- Vergleich der natürlichen Oxidation in Reinraumluft und in DI-Wasser,
- Vergleich der Oxiddicke von nach HF-Dip gespülten und ungespülten Wafern,
- Oxidentfernung in geringer konzentrierter Flusssäure.

Eine Übersicht über die einzelnen Prozeduren ist in der Tabelle 4.5 zusammengestellt.

Tabelle 4.5: Übersicht über die Reinigungs- und Oxidentfernungsprozeduren.

Wafer-Nr.	Reinigungsschritt	1. HF-Dip	2. – 5. HF-Dip	Bemerkungen
A-1	Caro 120 °C, 5 min	60 s, 1 %	60 s, 1 %	
A-2	Caro 120 °C, 5 min	120 s, 1 %	60 s, 1 %	1 x 10,5 h Spülen in DI-Wasser nach HF
A-3	Caro 120 °C, 5 min	30 s, 1 %		
B-1	SC1 80 °C, 5 min	60 s, 1 %		ohne Spülen in DI-Wasser nach HF
B-2	SC1 80 °C, 10 min	60 s, 1 %		
C-1	Caro 120 °C, 5 min	300 s, 0,1 %		
C-2	Caro 120 °C, 5 min	600 s, 0,1 %		

4.3.2 Temperaturkalibrierung der Anlage

4.3.2.1 Untersuchungen der Zuverlässigkeit der Temperaturmessung

Bei einem Kurzzeitprozess (RTP) sind Temperatur und Zeit die entscheidenden Prozessgrößen. Mit kürzeren Prozesszeiten kann bei gleichem thermischen Budget die Prozesstemperatur so weit erhöht werden, dass thermisch aktivierte Prozesse effektiver durchgeführt werden können. Da die thermisch aktivierten Prozesse in der Regel exponentiell von der Temperatur abhängen, können minimale Schwankungen der Temperatur zu großen Unterschieden beim Prozessergebnis führen.

Die Hauptschwierigkeit bei den sehr hohen Temperaturen und den durch die kurzen Zeiten bedingten schnellen Temperaturänderungen ist deren genaue Erfassung und Kontrolle über den gesamten Bereich des Wafers. Dazu müssen Verfahren eingesetzt werden, die es erlauben, die Temperatur des Wafers selbst zu ermitteln, ohne diese durch die Messung zu verändern. Diese muss dabei eine geringe thermische Trägheit aufweisen und mindestens so schnell auf Temperaturveränderungen reagieren, wie sie vom Prozess vorgegeben sind.

Für schnellste berührungsfreie Temperaturmessungen bieten sich optische Verfahren wie die pyrometrischen Messungen an. In der verwendeten Anlage stand jedoch eine solche Temperaturmessung nicht zu Verfügung. Ausgestattet war die Anlage mit einer Temperaturerfassung mittels PinTC. So war kritisch zu überprüfen, ob diese zwar einfache und weniger aufwendige Möglichkeit der Temperaturmessung bei den Dotierprozessen eine zuverlässige Echtzeit-Temperaturerfassung gestattet. Diese Temperaturbestimmung ist nicht nur für die Bewertung der Vorgänge bei der Dotierung erforderlich, sondern ist Grundlage für die Kontrolle und Reproduzierbarkeit der Prozesse, da über die Closed-Loop-Regelung Abweichungen des vorgegebenen Temperatur-Sollwerts von dem gemessenen aktuellen Istwert der Temperatur über die Veränderung der Regelgröße Lampenleistung minimiert werden.

Stabilität und Reproduzierbarkeit der Hochtemperaturprozesse

Im Laufe der Experimente zeigte sich, dass die Temperaturerfassung über den PinTC sehr störanfällig und teilweise unzuverlässig war. Aus diesem Grunde waren eingehendere Betrachtungen zur Temperaturerfassung erforderlich. Dazu wurden umfangreiche Untersuchungen zum Zusammenhang der Lampenleistung, einem unabhängig von der Temperaturerfassung aufgezeichneten Prozessparameter, und den Prozessergebnissen durchgeführt. Diese erfolgten im Hinblick auf die Stabilität und vor allem Reproduzierbarkeit der Hochtemperaturprozesse. Es wurden dazu verschiedene Prozesse verglichen, die mit gleichen Rezepten durchgeführt wurden, und Abweichungen von Prozessverlauf und -ergebnissen evaluiert.

Einsatz verschiedener Thermolemente

Es ergab sich im Laufe der Experimente, dass es zum wiederholten Ausfall des PinTCs der Anlage kam. Die Möglichkeit eines Austauschs des Thermolements des Original-PinTCs unter weiterer Verwendung des Quarzaufbaus war nicht möglich. Nach mehrmaligem Ersatz musste aus Kostengründen ein Eigenbau verwendet werden. Zwei parallel geführte und durch Zusammenschmelzen verbundene Quarzglasröhrchen dienten der Aufnahme jeweils eines Thermoschenkels des Thermolements, ein seitlich angebrachtes Quarzröhrchen diente als Öse zur Befestigung am Wafertray. Die Röhrchen zur Führung des Thermolements waren so geformt, dass sie an einem Ende als Waferauflage, entsprechend dem Pin des Originals, dienten. Dort befand sich auch die Messstelle. Dabei kamen, wie auch in Abschnitt 4.1 beschrieben, zwei Varianten zum Einsatz.

In einer ersten Ausführung war die Messstelle, das eigentliche Thermolement, frei über den Röhrchenenden, so dass der Wafer direkt darauf abgelegt wurde. Der Vorteil ist hier eine direkte, unmittelbare Messung am Wafer, was ein sehr schnelles Ansprechen und geringste thermische Trägheit versprach. Der Nachteil ist der direkte Kontakt der Thermolementmetalle zum Wafer, wobei mit gegenseitiger Kontamination zu rechnen ist, was beim Thermolement zu einer schnellen Alterung führen könnte.

In einer zweiten Ausführung wurde die Messstelle mit einer geringen Menge Keramikzement an den Röhrchenenden fixiert und somit gleichzeitig verkapselt. Verschiedene Prozesse unter Einsatz der verschiedenen Thermolementausführungen wurden verglichen.

4.3.2.2 Einsatz von mit Thermolementen bestückten Wafern

Eine quantitative Erfassung der Messfehler des Pin-Thermolements ist nur durch unabhängige direkte Messungen der Wafertemperatur möglich. Dafür wurde ein so genannter TC-Wafer (thermocouple wafer) bei der Ausführung von Testprozessen eingesetzt.

Der TC-Wafer besitzt ein in eine kleine Vertiefung auf der Oberfläche eingeklebtes Thermolement. Hierfür wurde ein Thermolement vom Typ S in der Wafermitte positioniert. Das Einlegen des TC-Wafers in die Kammer kann nur manuell erfolgen, Unterschiede der Waferposition gegenüber der Lage der Prozesswafer sind dabei unvermeidlich. Aus Fotos von der Waferlage in der Prozesskammer (wie in Abbildung 4.2 für Prozesswafer angegeben) konnten die genauen Auflagepunkte bestimmt werden. Sie sind in der Abbildung 4.5 im Vergleich mit denen der mittels Handler eingelegten Prozesswafer dargestellt.

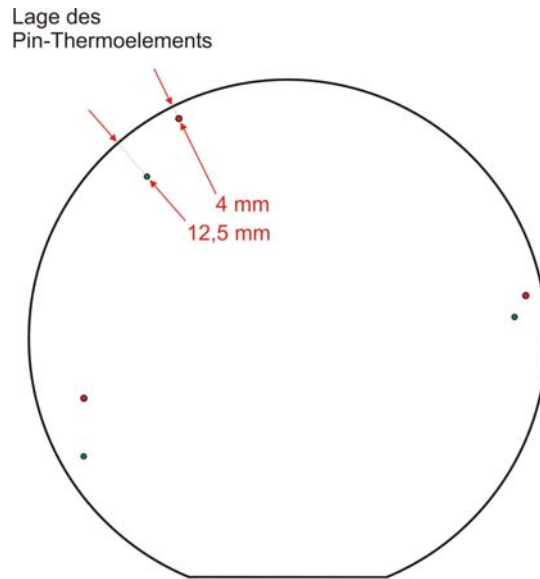


Abbildung 4.5: Lage der Auflagepunkte der Wafer in der Prozesskammer (Ansicht von unten). Die Lage des Pin-Thermoelements mit jeweiligem Abstand zum Wafer-Rand ist markiert.

Rote Punkte: Lage der Prozesswafer (mit Handler abgelegt),

Grüne Punkte: Lage des TC-Wafers.

Die Thermoelementdrähte sind innerhalb der Prozesskammer in Quarzröhrchen geführt und somit elektrisch isoliert. Die Durchführung nach außen erfolgt zwischen einer Vitondichtung und einem ebenen Kammerseitendeckel, wobei dieser eingeklemmte Bereich der Thermoelementdrähte mit einer dünnen selbstklebenden Kaptonfolie elektrisch isoliert ist. Das Thermoelement konnte an eine vorhandene Anlagenschnittstelle angeschlossen werden, die Messwerterfassung und -speicherung war damit über die Anlagensoftware möglich.

Die Temperaturkalibrierungen konnten erst nach den Dotierprozessen erfolgen. Die Testprozesse zur Bestimmung der Wafertemperatur in Abhängigkeit von der Temperatur des Pin-Thermoelements und von der benötigten Lampenleistung wurden in Anlehnung an die Prozessbedingungen bei der Dotierung gewählt und durchgeführt. Ein Beispiel für den Verlauf der Solltemperatur ist in der Abbildung 4.6 angegeben.

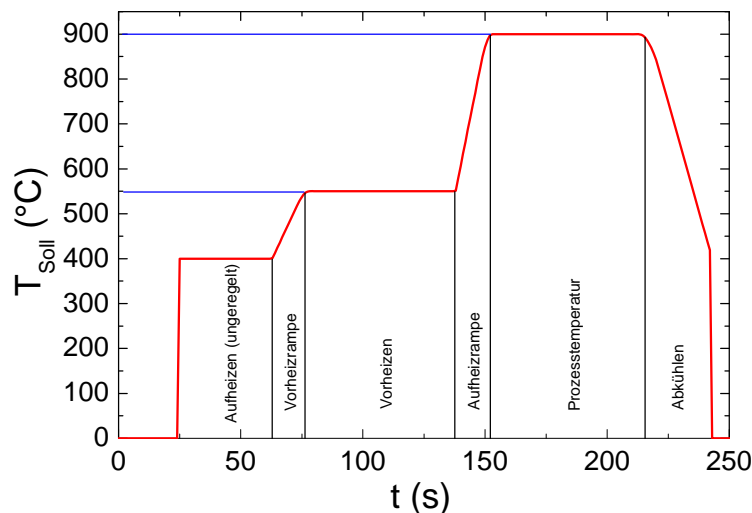


Abbildung 4.6: Beispiel des Verlaufs der Solltemperatur eines 900°C-Prozesses.

Die Testprozesse erfolgten auch bei stationärem Sauerstofffluss von 100 sccm, was zu einem Prozessdruck von ca. 5 mbar führte. Dies ist außerdem günstig, um ein Ätzen des Wafers durch SiO-Bildung bei Sauerstoffmangel zu vermeiden, und günstig für die Stabilität der verwendeten Thermolemente. Im Folgenden soll nur die Vorgehensweise der Temperaturbestimmung erläutert werden, die Ausführungen zu den eigentlichen Dotierprozessen erfolgen weiter unten.

4.3.2.3 Kurzzeitoxidation von Testwafern

Die sehr schnellen Kurzzeitprozesse laufen nicht im thermischen Gleichgewicht ab. Temperaturunterschiede im Silicium durch lokale Unterschiede der Strahlungsaufnahme können sich während der Dauer des Prozesses nicht ausgleichen. Daher ist die Kontrolle von Temperaturinhomogenitäten erforderlich und möglichst über eine Anpassung des Wärmeeintrags durch eine günstige Ausbildung des Strahlungsfeldes (durch konstruktive Maßnahmen wie Optimierung des Kammerinnendesigns, Verspiegelung der Oberflächen) und durch eine Einzelregelung von Teilbereichen der Strahlungsquelle als aktive Methode der Optimierung auszugleichen. Wie in Abschnitt 4.1 beschrieben, bestand beim verwendeten Reaktor die Strahlungsquelle aus 16 nebeneinander angeordneten Stablampen unterhalb des Wafers, die einzeln angesteuert werden konnten. Eine Anpassung der Strahlungsleistung war somit nur in Richtung der Lampenreihung, jedoch nicht innerhalb der Ausdehnung der einzelnen Lampen möglich.

Die Aufgabe der Homogenitätsuntersuchungen sollte nun darin bestehen, die Möglichkeiten der Anlage für die Erzielung gleichmäßiger Prozessergebnisse über die gesamte Ausdehnung der eingesetzten 150 mm-Wafer einzuschätzen, Ursachen für Inhomogenitäten zu bestimmen und die zur Verfügung stehenden Mittel einer Korrektur zu prüfen.

Als Monitorprozess wurden Kurzzeitoxidationen in der Kammer durchgeführt, von denen erwartet wurde, dass

- sie als einfache temperaturaktivierte Prozesse reproduzierbare Resultate erbringen,
- sie vor allem die Temperaturunterschiede auf dem Wafer als Schwankungen in der Verteilung der Oxidschichtdicken auf der Waferoberfläche eindeutig und empfindlich genug abbilden,
- diese nach dem Prozess nicht mehr veränderlich sind und dann mittels Ellipsometer ortsaufgelöst vermessen werden können.

Sollten noch andere Einflüsse als die Temperaturverteilung auf diese Schichtdickenschwankungen bestehen, wie Unterschiede in lokalen Druck- oder Strömungsverhältnissen, würden diese bei den Dotierprozessen auch auftreten, da die Oxidationsbedingungen entsprechend gewählt wurden. Außer durch ein geeignetes Kammerdesign sind solche Einflüsse jedoch kaum zu beeinflussen.

Der Nachteil von diesem „ex-situ-Monitoring“ ist, dass keine temporären Temperaturschwankungen zu ermitteln sind. Das Oxidationsergebnis stellt gewissermaßen die Summe aller Einflüsse dar. Für eine Einschätzung der Leistungsfähigkeit der Anlage sollte das jedoch keine Einschränkung darstellen, da dies gleichermaßen für die Ergebnisse der Dotierprozesse gilt.

Bei den Versuchen wurde darauf geachtet, gleiche Ausgangsbedingungen zu schaffen. Alle Wafer wurden vor ihrer Prozessierung durch die normale Eingangsreinigung von Verunreinigungen und der natürlichen Oxidschicht befreit. Unmittelbar nach der Reinigung

wurden die Waferhorden in die Beladeschleuse eingelegt, wobei die Flats entsprechend ausgerichtet wurden.

Zwei Methoden der Oxidation wurden untersucht, die auch den Prozessverhältnissen während der Dotierung entsprechen sollten:

- „statische“ Oxidation
Das Kammervakuumventil wurde geschlossen und Sauerstoff über eine definierte Zeit eingelassen. Danach wurden die Temperaturschritte ausgeführt. Nach dem Ende der Temperaturrampen wurde der Sauerstoff abgepumpt.
- „stationäre“ Oxidation
Das Kammervakuumventil blieb während des gesamten Prozesses geöffnet, und der Sauerstoff wurde über die Dauer des Temperaturschritts mit dem maximal möglichen Fluss von 100 sccm eingelassen.

Bei der statischen Oxidation wurde nach 180 s Sauerstoffeinlass ein Druck von ca. 40 mbar erreicht, bei der stationären Oxidation konnte mit geringster Öffnung des Regelventils ein stationärer Druck von 5 mbar erzielt werden. Die Temperaturmessung bei diesen Versuchen erfolgte mittels des Thermoelements mit der Keramikplatte.

Die Messung der Oxiddicke erfolgte nach den Prozessen in der Reinraumumgebung mittels Ellipsometer, wie in Abschnitt 5.1 beschrieben. Es wurden Messrezepte mit automatischen Messungen an festgelegten Waferpositionen durchgeführt („Map“ aus 529 oder 355 Messungen im gleichen Abstand auf einem betrachteten Bereich von 130 mm Durchmesser). Damit stand für eine orts aufgelöste Datenauswertung und -darstellung eine ausreichend hohe Datenmenge zur Verfügung. Einzelne offensichtliche Messfehler konnten erkannt und verworfen werden. Für die Auswertungen wurde ein einfaches Schichtenmodell verwendet (siehe Abbildung 4.9), das für die gewachsenen Oxide sehr gute Anpassungen lieferte.

4.3.3 Dotierexperimente

Alle Dotierexperimente erfolgten mit gereinigten Wafern. Wie die Untersuchungen zeigten, erwies sich die nasschemische Reinigung, wie im Kapitel 4.2.2 beschrieben, als ausreichende Vorbehandlung, sie erfolgte bei allen Wafern gleich. Die Dotierprozesse wurden ausschließlich über die Rezeptsteuerung ausgeführt, so dass jeweils ein gleicher Ablauf gewährleistet war und eine Aufzeichnung der Prozessdaten erfolgen konnte.

Die Abläufe und Prozessparametervariationen der Dotierexperimente werden im Zusammenhang mit der Diskussion der Messergebnisse im Detail vorgestellt. Mit den Erkenntnissen aus den Experimenten wurden die Prozesse variiert und optimiert. So sollen an dieser Stelle die Dotierprozesse nur kurz vorgestellt werden.

Einstufiger Dotierprozess

Der einstufige Dotierprozess ist ein reiner Gasphasenprozess. Der Hochtemperaturschritt erfolgt, während in der Gasatmosphäre das Dotiergas mit einem festgelegten Partialdruck vorliegt. Dazu wurden Experimente mit dem Dotiergas höherer Phosphorkonzentration (10 % in N₂) und geringer Konzentration (50 vpm in H₂) durchgeführt.

Zweistufiger Dotierprozess

Der Hauptteil der Untersuchungen zur Phosphordotierung bezieht sich auf einen zweistufigen Prozess. Zuerst erfolgte dabei eine Adsorption von Phosphor in der Dotiergasatmosphäre, wofür ausschließlich das gering konzentrierte Phosphin eingesetzt wurde. Danach erfolgte in einem kurzen Hochtemperaturprozess ohne weitere Zufuhr von Phosphor aus der Gasphase ein Eintreiben und Aktivieren der Dopanten.

4.4 Charakterisierung der dotierten Siliciumwafer

4.4.1 Untersuchungen zur Belegung der Oberfläche

4.4.1.1 Ellipsometrie

Die Ellipsometrie gehört zu den optischen Messverfahren. Die Grundlage ist dabei die Veränderung von elliptisch polarisiertem Licht aufgrund von Reflexion und Brechung an einer Probe. Es erfolgt in der Regel eine Analyse des von der Probe beeinflussten reflektierten Lichts im Vergleich zum einfallenden, jedoch sind auch Untersuchungen in Transmission möglich. Durch die Messung werden Informationen über die optischen Eigenschaften des gesamten, das Licht beeinflussenden Probensystems gewonnen. Mit diesen Informationen kann dann auf Eigenschaften der das System bildenden Bestandteile geschlossen werden, wie die Dicken von Schichten auf dem Substrat und deren optische Konstanten sowie die optischen Eigenschaften des Substrats selbst. Im Rahmen der vorliegenden Arbeit wurden ellipsometrische Untersuchungen angewendet, um Informationen zur Belegung der Siliciumoberfläche zu erhalten. Die vor dem Dotierprozess vorliegende Oberfläche sollte möglichst genau charakterisiert werden, um im Vergleich von unbehandelten mit dotierten Proben Oberflächenveränderungen durch den Dotierprozess zu ermitteln. Daher soll an dieser Stelle auf wichtige Aspekte bei der ellipsometrischen Untersuchung dünnster Schichten auf Silicium näher eingegangen werden. Eine entsprechende weitergehende grundlegende Beschreibung ist z.B. in [Tompkins '99] zu finden.

Grundlagen

Infolge der Reflexionen und Brechungen am Probensystem ergeben sich Veränderungen des Polarisationszustands der resultierenden reflektierten elektromagnetischen Welle gegenüber der einfallenden. Dies kann durch eine Ermittlung von Änderungen von Phase und Amplitude vom innerhalb, d.h. parallel (p-), und senkrecht (s-) zur Einfallsebene polarisierten Anteil der Welle bestimmt werden. Hierbei beziehen sich diese Angaben auf den elektrischen Feldvektor. Der resultierende Phasenunterschied aufgrund der in der Regel unterschiedlichen Reflexion der einzelnen Anteile wird als Δ bezeichnet. Wenn δ_1 der Phasenunterschied zwischen dem p- und s-Anteil der einfallenden Welle und δ_2 der entsprechende Phasenunterschied der reflektierten Welle ist, dann ergibt sich für Δ :

$$\Delta = \delta_1 - \delta_2, \quad (33)$$

wobei Δ von -180° bis $+180^\circ$ oder von 0° bis 360° angegeben werden kann.

Darüber hinaus hat die resultierende Reflexion an der Probenoberfläche eine für s- und p-Anteile im allgemeinen unterschiedliche Reduzierung der Amplituden zur Folge. Der totale

Reflexionskoeffizient des jeweiligen Anteils ist definiert als das Verhältnis der jeweiligen Amplituden des reflektierten und einfallenden Polarisationsanteils und wird mit R_s bzw. R_p bezeichnet. In der Regel sind R_s und R_p komplexe Größen. Die Beträge $|R_s|$ und $|R_p|$ sind damit das Maß der jeweiligen Amplitudenabnahmen. Das Verhältnis dieser beiden Beträge kann geeignet über die Angabe eines Winkels Ψ ausgedrückt werden,

$$\tan \Psi = \frac{|R_p|}{|R_s|}. \quad (34)$$

Die Größe Ψ ist also der Winkel, dessen Tangens das Verhältnis der Beträge der totalen Reflexionskoeffizienten von s- und p-polarisiertem Anteil darstellt, sein Bereich erstreckt sich von 0° bis 90° .

Das komplexe Verhältnis der totalen Reflexionskoeffizienten, das die Beeinflussung der elektromagnetischen Welle durch das Probensystem beschreibt,

$$\rho = \frac{R_p}{R_s}, \quad (35)$$

kann nun ausgedrückt werden als

$$\rho = \tan \Psi e^{i\Delta}. \quad (36)$$

Dies ist die Grundgleichung der Ellipsometrie. Der Betrag von ρ ist somit in $\tan \Psi$ enthalten und die Informationen zur Phase in der komplexen e -Funktion mit Δ . Die Informationen über die Veränderungen der Lichtwelle aufgrund von Reflexion können somit durch die Messung der Winkelgrößen Ψ und Δ zugänglich gemacht werden. Die Informationen über das Probensystem, das diese Änderungen bewirkt hat, wie die Werte der Brechungsindizes n , Extinktionskoeffizienten k und auch Schichtdicken d der verschiedenen involvierten Materialien sind in den totalen Reflexionskoeffizienten und damit in ρ enthalten.

Zu beachten ist, dass Ψ und Δ die eigentlichen Messgrößen der Ellipsometrie sind, die nach geeigneter vorheriger Kalibrierung aus der Kenntnis der Winkelstellungen der die Polarisation beeinflussenden optischen Baugruppen im Zusammenhang mit Intensitätsmessungen des reflektierten Lichts ermittelt werden können. Die im Allgemeinen interessierenden Informationen zu den abgeleiteten Größen n , k und d hängen jedoch entscheidend von der Gültigkeit des gewählten Modells ab. Bei genauer Kenntnis der optischen Eigenschaften des Probensystems können mit Hilfe der komplexen Theorie der Lichtreflexion die Intensitäten und Polarisationswinkel für das entsprechende System in Abhängigkeit von Wellenlänge und Einfallswinkel exakt berechnet werden. Umgekehrt ist jedoch nach Messung der Intensitäten und Winkel durch das Ellipsometer eine eindeutige Berechnung des optischen Systems nicht möglich oder zumindest äußerst kompliziert. Daher werden im Allgemeinen mittels der Software unter Vorgabe von Modellsystemen, die dem zu vermessenden möglichst nahe kommen sollen, ausgewählte Modellparameter so lange variiert, bis eine minimale Abweichung der berechneten Werte des Modells von den gemessenen gefunden wird.

Modelle für dünne Oxide auf Silicium

Die Beschreibung der optischen Eigenschaften des Probensystems erfolgt über geeignete Modelle der Materialabfolge, die das Licht beeinflusst. Für die einzelnen Materialien sind die optischen Parameter n und k (oder die dielektrischen Konstanten ε_1 und ε_2) in Abhängigkeit

von der für die Untersuchung benutzten Wellenlänge bekannt oder zumindest als Anfangsparameter für die nachfolgende Modellierung gegeben.

In der für die Auswertung der Ellipsometermessungen verwendeten Software SpectraRay der Firma SENTECH GmbH ist eine Materialdatenbank vorhanden, in der die für die Modellierung benötigten Parameter abgelegt sind. Für die Modellierung des Siliciumsubstrats wurden tabellarische Werte von n und k von Jellison verwendet, wie in der folgenden Abbildung 4.7 dargestellt.

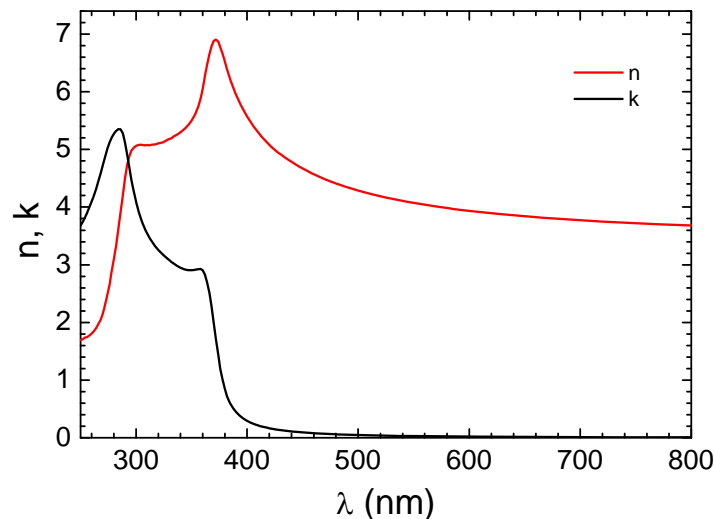


Abbildung 4.7: Brechungsindex n und Extinktionskoeffizient k von (100)-Silicium in Abhängigkeit von der Wellenlänge nach Jellison.

Für die meisten dielektrischen Schichten liegt in dem betrachteten Wellenlängenbereich von 250 nm bis 800 nm normale Dispersion vor, das heißt, der Brechungsindex fällt mit steigender Wellenlänge. Solch ein Verlauf kann gut über eine Polynomfunktion beschrieben werden. Dafür erweist sich die so genannte Cauchy-Funktion als günstig,

$$n(\lambda) = n_0 + n_1 \frac{10^2}{\lambda^2} + n_2 \frac{10^7}{\lambda^4}. \quad (37)$$

Als Standardbeschreibung einer Oxidschicht auf Silicium wurde im Folgenden das Cauchy-Modell SiO_2 (thermisch) mit den Parametern

$$\begin{aligned} n_1 &= 1,452, \\ n_2 &= 36, \\ n_3 &= 0 \end{aligned}$$

verwendet, der zugehörige Verlauf der Dispersionsfunktion ist in Abbildung 4.8 dargestellt. Auch für $k(\lambda)$ lässt sich eine entsprechende Polynomfunktion definieren, in den meisten Fällen, so auch für die betrachteten dünnen Oxide, ist k jedoch null im beschriebenen Spektralbereich. Die Verwendung von Formeln zur Beschreibung der Dispersion erweist sich insbesondere bei der Bestimmung von unbekanntem optischen Parametern n und k über die Parametervariation günstiger gegenüber der tabellarischen Form, da dann eine Variation von nur wenigen Formelvariablen erfolgen kann.

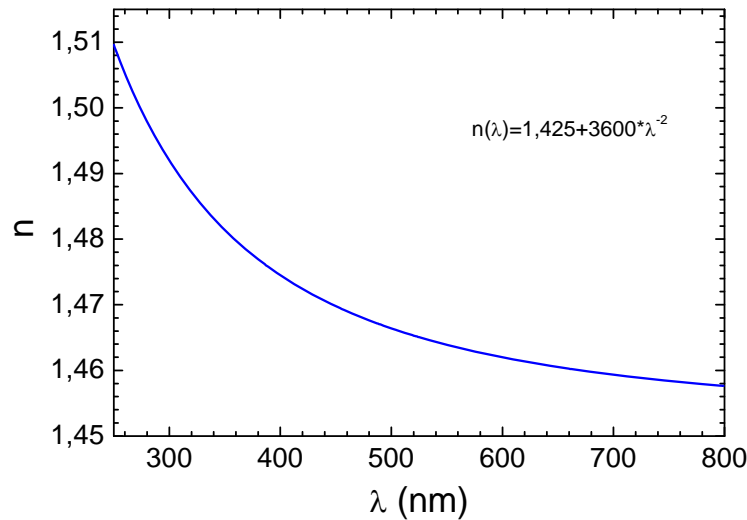


Abbildung 4.8: Brechungsindex n von Siliciumdioxid in Abhängigkeit von der Wellenlänge als Cauchy-Funktion in dem Bereich der normalen Dispersion (Cauchy-SiO₂ (thermisch)).

Das Modell, das nun das gesamte Probensystem beschreibt, wird aus all den Materialien zusammengesetzt, die die elektromagnetische Welle durchdringt. In der folgenden Abbildung ist ein Modell für die Beschreibung einer Oxidbedeckung angegeben.

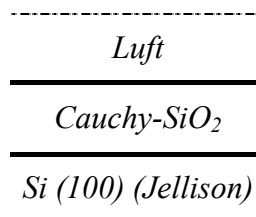


Abbildung 4.9: Schichtmodell für die Ermittlung der Oxidschichtdicken.

Die Luftschicht geht in diesem Modell durch die konstanten Werte $n = 1$ und $k = 0$ ein.

Soll nun die Oxiddicke aus den ellipsometrischen Messungen von Ψ und Δ ermittelt werden, wird in diesem Modell allein der Wert der Dicke der Cauchy-SiO₂-Schicht so lange variiert, bis eine minimale Abweichung zwischen gemessenen Werten und den für das Modell berechneten Werten gefunden wird. Ein Maß hierfür ist der mittlere quadratische Fehler (MSE - mean square error). Bei einem Datensatz aus N Ψ - und Δ -Werten wird er wie folgt bestimmt:

$$MSE = \frac{1}{N} \sqrt{\sum_{i=1}^N [(\Psi_i^m - \Psi_i^{th})^2 + (\Delta_i^m - \Delta_i^{th})^2]}, \quad (38)$$

wobei die Indizes ^m für den gemessenen Wert und th für den entsprechenden durch die Variation ermittelten theoretischen Wert stehen.

Analyse von dünnen Schichten auf Silicium

Wenn die optischen Konstanten und das Modell bekannt sind und somit auch die tatsächlich vorliegende Struktur repräsentieren, ist die Ellipsometrie geeignet, die Dicke von sehr dünnen

Oxiden auf Silicium zu messen. Die messbaren Einflüsse der dünnen Schicht machen sich dabei fast ausschließlich im Δ -Spektrum (den wellenlängenabhängigen Δ -Werten) bemerkbar, während Ψ für geringe Oxiddicken im gleichen Spektrum fast unabhängig von der Dicke ist. In der Abbildung 4.10 ist dies für eine Oxidbedeckung von 0 bis 5 nm und 70° Einfallswinkel dargestellt¹.

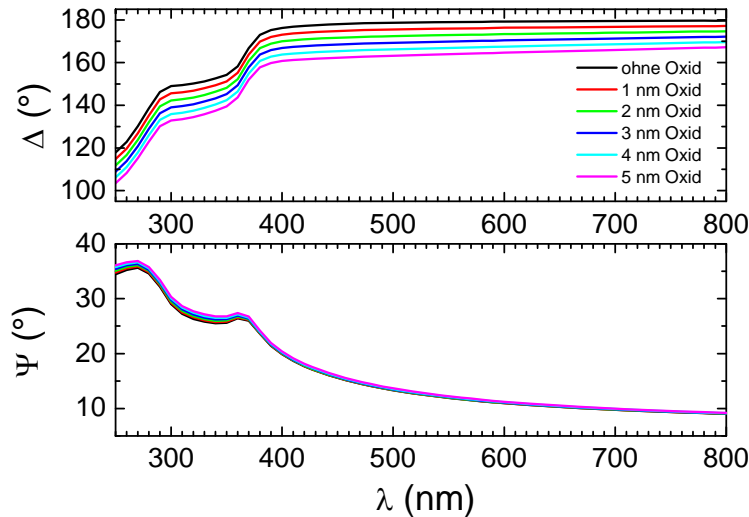


Abbildung 4.10: Berechnete Spektren von Δ und Ψ für 0 bis 5 nm Oxiddicke auf Silicium und 70° Einfallswinkel.

Durch die dünne Schicht wird die Intensität des reflektierten Lichts gegenüber einer Reflexion am reinen Silicium ohne Schicht kaum verändert, jedoch ändert sich der Phasenunterschied zwischen s- und p-Anteil signifikant und messbar. In der Abbildung 4.11 sind für einen Einfallswinkel von 70° die berechneten Werte von Δ und Ψ für zwei Wellenlängen in Abhängigkeit von der Schichtdicke dargestellt. Der wellenlängenabhängige Brechungsindex der Oxidschicht wurde hierbei mit einer Cauchy-Funktion aus Abbildung 4.8 beschrieben.

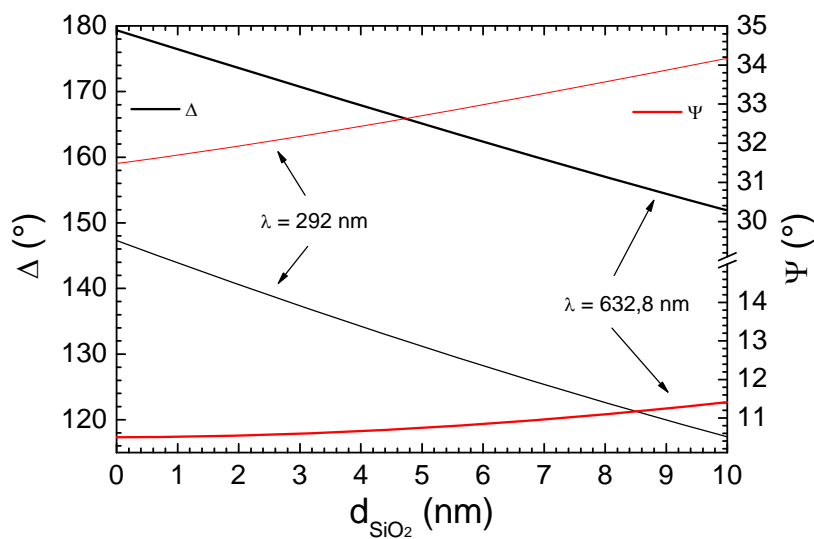


Abbildung 4.11: Veränderung von Δ und Ψ in Abhängigkeit von der Oxiddicke für zwei verschiedene Wellenlängen (berechnet).

¹ Diese und die folgenden Spektren wurden mit der Software SpectraRay der Firma SENTECH GmbH berechnet. Wenn nicht anders angegeben, wurde das Modell aus Abbildung 4.9 verwendet.

Die Änderung von Δ ist in dem dargestellten Schichtdickenbereich deutlich ausgeprägter als die von Ψ , wobei durch die Verwendung von kürzeren Wellenlängen eine Erhöhung der Empfindlichkeit auch von Ψ zu verzeichnen ist. Daher sollte für die Messung dünnster Schichten möglichst im UV-Bereich gearbeitet werden. Bei Vorhandensein einer 0,1 nm dicken Schicht auf Silicium würden sich gegenüber der unbedeckten Oberfläche die in der Tabelle 4.6 angegebenen Änderungen der Parameter ergeben.

Tabelle 4.6: Änderung von Δ und Ψ durch eine 0,1 nm dicke Oxidschicht gegenüber unbedeckter Siliciumoberfläche für zwei verschiedene Wellenlängen bei einem Einfallswinkel von 70° .

λ	n_{SiO_2}	Änderung Δ	Änderung Ψ
292 nm	1,461	0,29°	0,0005°
632,8 nm	1,494	0,34°	0,027°

Die Änderung von Δ ist hierbei größer als die Messungenauigkeiten gewöhnlicher Ellipsometer, so dass sich damit Monolagenbedeckungen bestimmen lassen sollten.

Bei Schichten, die nur wenige nm dick sind, ist die Änderung von Δ eine aussagekräftigere Größe als die absoluten Werte von Δ für das System mit und ohne Schicht. Dies zeigt sich bei Experimenten von in-situ Messungen von Filmwachstum bei unveränderter Messposition, so dass die Änderung von Δ genauer zu bestimmen ist als der Absolutwert selbst. Ein Beispiel hierfür ist ein klassisches Experiment von Archer, der die Chemisorption von Sauerstoff im UHV auf gespaltenem Silicium untersuchte [Archer '65]. Dabei zeigte sich eine eindeutige Veränderung von Δ mit der Dauer der Sauerstoffeinwirkung, wobei für ca. $0,7^\circ$ eine Sättigung beobachtet wurde. Dies entsprach mit ca. 0,25 nm Schichtdicke der Bedeckung einer Monolage. Mit solcher Messung waren somit auch Bruchteile von Monolagen ellipsometrisch messbar.

Da für sehr dünne Schichten (kleiner als 10 nm) die Werte für Ψ relativ unabhängig von der Schichtdicke sind, ist eine Bestimmung von optischen Konstanten selbst bei genauer Kenntnis der Schichtdicke schwierig bis unmöglich. So ist es, selbst bei Anwendung von spektroskopischer Ellipsometrie, nicht ratsam, für Schichten dünner als 10 nm die optischen Konstanten zu bestimmen. Andererseits genügt es für eine Dickenbestimmung solcher Schichten, ein optisches Modell zu benutzen, das dem tatsächlichen nur in etwa entspricht. Die so bestimmte Dicke ist dann eine effektive Dicke gemäß dem verwendeten Modell, wobei die Dickenänderung wie oben ausgeführt exakter zu bestimmen ist.

Eine sichere Unterscheidung von unterschiedlichen Schichtzusammensetzungen dünnster dielektrischer Schichten allein anhand gemessener Ψ - Δ -Spektren ist jedoch kaum möglich, wie die Abbildung 4.12 und die Tabelle 4.7 verdeutlichen.

Tabelle 4.7: Ermittelte Schichtdicken und -fehler aus Auswertung des Spektrums von Abbildung 4.12 unter Annahme der angegebenen Schichten auf reinem Silicium (Cauchy-Schichtmodelle aus SpectraRay).

Schicht	n (bei 633 nm)	d in nm	Δd in nm
SiO ₂ (therm.)	1,461	0,557	0,018
H ₂ O	1,332	0,664	0,021
Si ₃ N ₄	2,018	0,444	0,014
TiO ₂	2,386	0,460	0,014

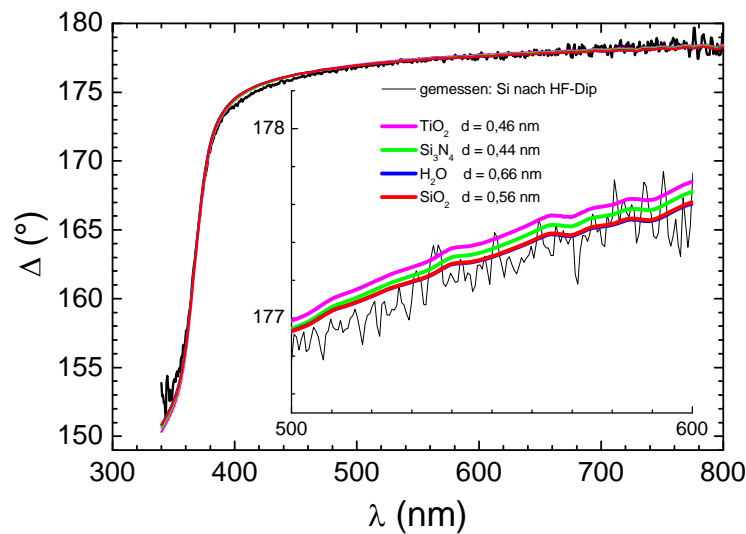


Abbildung 4.12: Gemessenes Δ -Spektrum von Silicium nach einem HF-Dip zusammen mit entsprechenden besten Approximationen der Dicke von vier verschiedenen dielektrischen Schichten auf reiner Siliciumoberfläche (Ψ -Spektren nicht mit dargestellt, da praktisch identisch).

Man kann erkennen, dass sich die Spektren, die sich aus der Berechnung von Schichtmodellen mit unterschiedlichen Materialien ergeben, fast identisch sind und auch gut zum gemessenen Spektrum passen. Die aus der Berechnung ermittelten Schichtdicken und deren Fehler (ermittelt u.a. aus der Größe der geringsten Abweichung von berechnetem und gemessenem Spektrum) sind in der Tabelle dargestellt.

Der Grund für diese Schwierigkeit bei der Unterscheidung dünner Schichten ist, dass die dielektrischen Schichten in dem angegebenen Wellenlängenbereich zwar unterschiedliche Beträge des Brechungsindex aufweisen, doch ist die Form der Dispersionskurve n über λ ähnlich (und approximierbar durch eine Cauchy-Funktion). Ebenso ist es auch äußerst schwierig, bei Schichtstapeln dünner Dielektrika die Dicke der Einzelschichten eindeutig zu ermitteln.

Rauigkeit

Auch Oberflächen- und Grenzflächenrauigkeiten können optische Messungen beeinflussen. Bei einer rauen Fläche zwischen zwei Medien, wie in Abbildung 4.13 dargestellt, können drei Bereiche unterschieden werden. Die Bereiche 1 und 2 mit den bekannten optischen Konstanten für die jeweiligen Medien und ein Zwischenbereich 3 mit unbestimmten optischen Eigenschaften, dessen Ausdehnung durch die Rauigkeit bestimmt ist. Dieser Bereich kann als Schicht interpretiert werden, mit optischen Eigenschaften, die zwischen denen von Schicht 1 und 2 liegen. Für Ausdehnungen weit geringer als die eingesetzte Wellenlänge kann dies angenommen werden. Entsprechend kann diese Rauigkeit als zusätzliche Schicht in das Modell der Probenstruktur eingebunden werden.

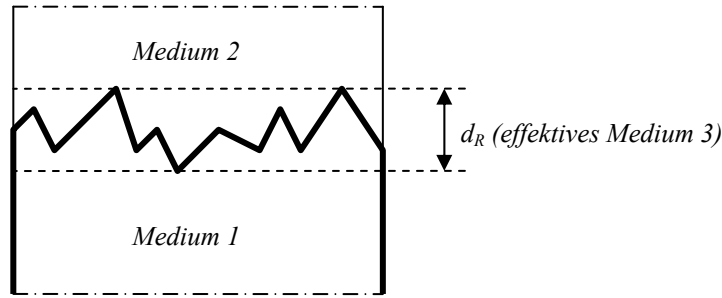


Abbildung 4.13: Modellierung einer rauen Grenzfläche durch eine angenommene Schicht der Dicke d_R .

Ein weiterer Einfluss der rauen Fläche ist die Streuung des einfallenden Lichts an den von der betrachteten Reflexionsebene abweichenden Flächenbereichen. Dieser Effekt führt zu einem wellenlängenabhängigen Intensitätsverlust des reflektierten Lichts für den betrachteten Ausfallswinkel. Für ellipsometrische Untersuchungen ist dieser Effekt in der Regel nicht bedeutend, es sei denn, die Rauigkeit ist so groß, dass dadurch Depolarisationseffekte auftreten, die kaum zu modellieren sind.

Für geringe Rauigkeiten an Siliciumoberflächen ist es kaum möglich, diese von einem dünnen Oxid zu unterscheiden. Wie bei den dünnen Schichten oben beschrieben, ist es somit nicht eindeutig möglich, eine Rauigkeit oder eine Oxiddicke zweifelsfrei zu bestimmen. Andererseits kann man die Rauigkeit auch durch die Annahme einer Oxidbedeckung modellieren. Eine bessere Beschreibung der Oberflächen- und der Grenzflächenrauigkeit liefert eine Approximation durch ein so genanntes effektives Medium (EMA - effective medium approximation). Die Rauigkeitsschicht wird hierbei als Medium mit eigenen festgelegten optischen Eigenschaften betrachtet. Das einfachste Modell ist eine lineare Interpolation der beiden begrenzenden Medien 1 und 2,

$$\tilde{\varepsilon} = f_1 \tilde{\varepsilon}_1 + f_2 \tilde{\varepsilon}_2, \quad \text{mit } f_1 + f_2 = 1, \quad (39)$$

wobei $\tilde{\varepsilon}$ die effektive komplexe dielektrische Funktion des Rauigkeitsmediums, $\tilde{\varepsilon}_1$ und $\tilde{\varepsilon}_2$ die komplexen dielektrischen Funktionen der angrenzenden Medien und f_1 und f_2 deren Volumenanteile im effektiven Medium sind.

Noch genauer als durch eine lineare Approximation lassen sich solche effektiven Medien jedoch durch die Annahme einer Einmischung eines Mediums 1 in ein Substratmedium 2 beschreiben. Man geht dann in der Regel von der vereinfachten Annahme von kugelförmigen Einschlüssen und Dipol-Wechselwirkungen aus. Gewöhnlich wird hier das Brüggemann-Modell, das auch bei in etwa gleichen Anteilen von Einschlüssen und Substrat funktioniert, benutzt. Für zwei Medien ergibt sich die folgende Bedingung,

$$0 = f_1 \frac{\tilde{\varepsilon}_1 - \tilde{\varepsilon}}{\tilde{\varepsilon}_1 + 2\tilde{\varepsilon}} + f_2 \frac{\tilde{\varepsilon}_2 - \tilde{\varepsilon}}{\tilde{\varepsilon}_2 + 2\tilde{\varepsilon}}. \quad (40)$$

Verwendetes Ellipsometer

Die Proben wurden mit einem Ellipsometer SE 850 der Firma SENTECH GmbH vermessen. Die Berechnungen wurden mit der Ellipsometer-Software SpectraRay 2 durchgeführt. Bei dem Gerät handelt es sich um ein spektroskopisches Ellipsometer. Verlässliche Messungen ließen sich im Wellenlängenbereich von etwa 350 nm bis 850 nm (5,54 eV bis 1,46 eV)

durchführen. Das Licht aus einer Xenon-Hochdruck-Bogenentladungslampe wurde über Lichtleitkabel in den Polarisatorarm geführt. Der Polarisator war nach einer Kalibrierung auf einer festgelegten Position. Nach Reflexion von der Probe, die auf einem Vakuumchuck auf einem bis zu 150 mm verfahrbaren x-y-Tisch an Luft eingespannt war, trat das Licht in den mit einem stufenweise drehbaren Analysator ausgerüsteten Analysatorarm ein und wurde über ein weiteres Lichtleitkabel zum Monochromator mit CCD-Zeilendetektor geleitet. Mit diesem Aufbau konnte eine Multiplex-Analyse im Step-Scan-Analysatorprinzip durchgeführt werden, was schnelle Messungen in Zeiten unter 30 s für ein komplettes Spektrum ermöglichte. Außerdem waren Messungen mit einem Cut-Off-Filter zum Unterdrücken von höheren Ordnungen im längerwelligen Bereich möglich und ein präziser Retarder für Messungen im Wellenlängenbereich oberhalb von 340 nm vorhanden. Dieser Retarder gestattete sehr genaue Messungen des Δ -Wertes auch nahe 180° , was bei dünnen Schichten auf Silicium erforderlich war. Die Messungen wurden bei festem Einfallswinkel von 70° durchgeführt und erwiesen sich in dem gewählten Wellenlängenbereich als sehr gut reproduzierbar.

4.4.1.2 Auger-Elektronenspektroskopie (AES)

Einige Untersuchungen zur Ermittlung der chemischen Zusammensetzung der Oberfläche konnten im Mikrostrukturzentrum der Universität in der Fakultät für Naturwissenschaften in einem AES-Modul einer Oberflächenanalyseanlage ESCA 5600 der Firma PERKIN ELMER durchgeführt werden. Dieses Verfahren soll an dieser Stelle jedoch nur kurz vorgestellt werden, da es nicht im Zentrum der Untersuchungen stand.

Die zu untersuchende Probe wird mit einem Elektronenstrahl angeregt, durch welchen Elektronen aus den inneren Schalen der Oberflächenatome herausgeschlagen werden. Diese Atome gelangen wieder in den Grundzustand, indem Elektronen aus höheren Schalen die frei gewordenen Plätze unter Energieabgabe wieder auffüllen. Ein Teil der freiwerdenden Energie führt wiederum zu einer Anregung von Elektronen auf äußeren Schalen, die dadurch den Atomverband verlassen können. Diese so genannten Auger-Elektronen können in einem Halbkugelanalysator energie- und richtungsabhängig detektiert werden. Ein verwandtes Verfahren, das auch auf einer Analyse von emittierten Elektronen beruht und daher oft in dieselbe Anlage zur Oberflächencharakterisierung integriert ist, wird als Röntgen-Photoelektronenspektroskopie (XPS) bezeichnet. Hier erfolgt jedoch die Anregung mit monochromatischer Röntgenstrahlung, und dadurch aus den Oberflächenatomen freigesetzte Elektronen werden detektiert. Diese kommen aus weiter innen liegenden Schalen als die Auger-Elektronen und sind energiereicher. Aus der kinetischen Energie der emittierten Elektronen lassen sich nun Rückschlüsse auf Art und Bindungsverhältnisse der oberflächennahen Elemente ziehen. Außer Wasserstoff und Helium sind alle Elemente detektierbar. Der Vorteil beim AES ist nun, dass durch die bessere Fokussierbarkeit des anregenden Elektronenstrahls eine höhere Ortsauflösung möglich ist. Außerdem können aufgrund ihrer geringen Energie nur die Auger-Elektronen von Oberflächenatomen der obersten Monolagen das Substrat verlassen, die von weiter innen liegenden Atomen werden vom Substrat absorbiert. Die energiereicheren Photoelektronen der Röntgenanregung können hingegen auch aus tieferen Schichten stammen. Daher ist auch die Tiefenauflösung beim AES höher. Nachteilig ist hier jedoch die deutlich geringere Elektronenausbeute und damit auch die geringere Messempfindlichkeit. Um Tiefenprofile aufnehmen zu können, wird die Oberfläche nach einer Messung durch Sputtern mit Argonionen schichtweise abgetragen. Mit diesem Verfahren wird die chemische Konzentration der in der Probe enthaltenen Stoffe prozentual ermittelt.

Für die AES-Untersuchungen mussten die Wafer zerteilt und aus der Reinraumumgebung ausgeschleust werden. Die Siliciumproben wurden mit einer Beschleunigungsspannung von 10 kV und einem Strahlstrom von 50 nA untersucht.

4.4.2 Ermittlung von Siliciumeigenschaften nach der Dotierung

4.4.2.1 Messung des Flächenwiderstands mittels Vierspitzenmethode

Ein Mittel zur Charakterisierung von Dotierungen in Silicium ist die Bestimmung des Flächenwiderstands; in der Technologie-Roadmap (siehe Kapitel 2.2.3) stellt diese Größe ein wichtiges Kriterium für die Schaffung von pn-Übergängen dar. Eine Flächenwiderstandsmessung ist relativ einfach auszuführen, kann als absolutes Messverfahren ohne Einsatz von Kalibrierproben erfolgen und wird daher auch insbesondere in Abhängigkeit von der Position auf der Waferoberfläche für Homogenitätsuntersuchungen zur Bestimmung von Widerstandsverteilungen routinemäßig durchgeführt.

Definition des Flächenwiderstands

Der spezifische Widerstand ρ als wichtiger elektrischer Materialparameter kann errechnet werden aus

$$\rho = \frac{1}{q(n\mu_n + p\mu_p)}, \quad (41)$$

wobei q die Elementarladung, μ_n und μ_p die Elektronen- bzw. Löcherbeweglichkeit sowie die insbesondere durch die Dotierung zu beeinflussenden Größen n und p die entsprechenden freien Ladungsträgerdichten sind. Oft sind diese Werte jedoch schwer zu bestimmen und daher unbekannt, so dass der spezifische Widerstand durch Messungen ermittelt werden muss. Die Charakterisierung von dünnen Schichten oder dünnen Materialbereichen, wie Dotierungsbereichen, erfolgt nun in der Regel eher durch die Angabe des Flächenwiderstands R_S als durch ρ oder den elektrischen Widerstand R einer bestimmten Geometrie. Dies ist günstig, da in R_S auch die Dicke der Schicht mit eingeht, die dann nicht explizit bekannt sein muss. Aus

$$R = \rho \frac{l}{b \cdot d}, \quad (42)$$

mit der Länge l , der Breite b und der Dicke d der Struktur, kann für $l = b$ geschrieben werden:

$$R_s = \frac{\rho}{d}. \quad (43)$$

Dies ist die Definition des Flächenwiderstands, der oft auch einfach als Schichtwiderstand bezeichnet wird. Seine Einheit ist Ω , sie wird gewöhnlich jedoch, so auch in dieser Arbeit, als Ω/square (Ω/sq oder Ω/\square) angegeben, da der Flächenwiderstand aufgrund der Definition von $l = b$ den Widerstand eines beliebigen quadratischen Flächenelements der so charakterisierten Struktur angibt. Somit ist es möglich, allein aus der Kenntnis von der Form und Fläche von Oberflächenschichten mit homogenem R_S deren elektrischen Widerstand lediglich durch Reihen- und Parallelschaltung von die Gesamtfläche ausfüllenden quadratischen Flächenelementen zu bestimmen.

Messung mittels Vierspitzenmethode

Der Flächenwiderstand eines Materials kann mittels Vierspitzenmessungen bestimmt werden. Die Widerstandsbestimmung erfolgt dabei über Strom- und Spannungsmessungen an den Spitzen. Vier Spitzen werden hierfür eingesetzt, da an den stromführenden Spitzen zusätzliche Widerstände, wie die Widerstände der Spitzen selbst, Kontaktwiderstände und Ausbreitungswiderstände von bzw. zu der jeweiligen Spitze auftreten, die mit dem interessierenden Flächenwiderstand in Reihe geschaltet sind und die nicht durch anderweitige Messungen ermittelt oder aber errechnet werden könnten. Eine separate Spannungsmessung an zwei anderen Spitzen kann hingegen praktisch stromlos erfolgen, so dass diese parasitären Widerstände dort nicht ins Gewicht fallen und die so ermittelte Potentialdifferenz eindeutig aufgrund des gesuchten Widerstands für den eingepprägten Stromfluss vorliegt. Diese Vierspitzenkonfiguration wird auch oft als Kelvinanordnung bezeichnet. Eine ausführliche Beschreibung mit weiterführenden Literaturangaben ist in [Schroder '98] zu finden.

Für die Messung wird ein Messkopf mit vier federnd gelagerten in der Regel in Reihe angeordneten Metallspitzen mit einer definierten, reproduzierbaren Kraft (meist 0,5 N bis 2,0 N) auf das zu untersuchende Material aufgedrückt (siehe Abbildung 4.14). Der Speisestrom sollte dabei 10 mA nicht übersteigen, um eine Erwärmung der Halbleiterprobe und damit Eigenleitungseffekte, die Messfehler verursachen, zu vermeiden.

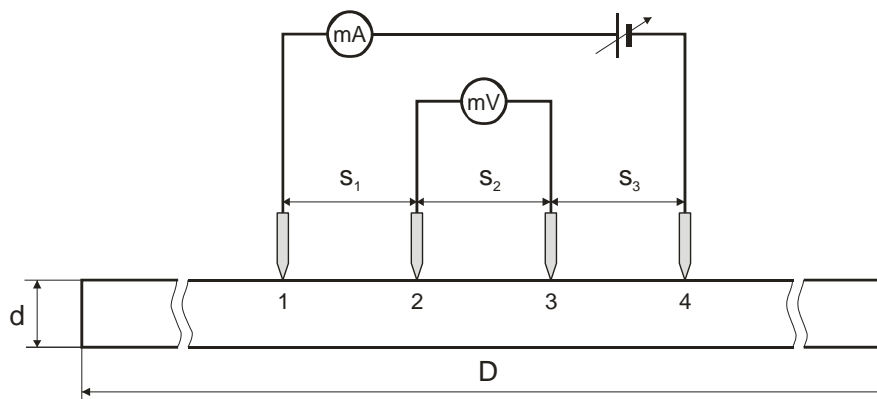


Abbildung 4.14: Prinzipaufbau der Vierspitzen-Messanordnung.

Für die Bestimmung des Widerstands aus den Strom- und Spannungsmessungen wird zunächst davon ausgegangen, dass es sich um ein zu untersuchendes Material von unendlicher Fläche und Dicke handelt. Das Potential in der Entfernung r zu einer stromführenden Spitze ist

$$\varphi = \frac{\rho \cdot I}{2\pi \cdot r}. \quad (44)$$

Im Abstand r_1 und r_4 von den äußeren stromführenden Spitzen 1 und 4 (vgl. Abbildung 4.14) ergibt sich

$$\varphi = \frac{\rho \cdot I}{2\pi} \left(\frac{1}{r_1} - \frac{1}{r_4} \right). \quad (45)$$

Für die Spannungsmessspitzen 2 und 3 folgt daraus unter Verwendung der Abstände s_1 , s_2 und s_3 ,

$$\varphi_{2,3} = \varphi_2 - \varphi_3 = \frac{\rho \cdot I}{2\pi} \left(\frac{1}{s_1} + \frac{1}{s_3} - \frac{1}{s_2 + s_3} - \frac{1}{s_1 + s_2} \right). \quad (46)$$

Die Potentialdifferenz $\varphi_{2,3}$ stellt die gemessene Spannung U dar, und nach Umstellen erhält man den spezifischen Widerstand an der Messstelle,

$$\rho = \frac{2\pi \cdot U}{I} \left(\frac{1}{s_1} + \frac{1}{s_3} - \frac{1}{s_2 + s_3} - \frac{1}{s_1 + s_2} \right)^{-1}. \quad (47)$$

Im Allgemeinen haben die Spitzen einen konstanten Abstand. Mit $s_1 = s_2 = s_3 = s$ ergibt sich

$$\rho = 2\pi s \frac{U}{I}, \quad (48)$$

als Gleichung zur Ermittlung des spezifischen Widerstands einer unendlich ausgedehnten Probe aus den Spannungs- und Strommessungen.

Korrekturfaktoren

Da reale Proben den idealen Anforderungen nicht genügen, werden drei Korrekturfaktoren eingeführt:

$$\rho = 2\pi s \frac{U}{I} \cdot K_1 \cdot K_2 \cdot K_3, \quad (49)$$

wobei

- K_1 den Einfluss der Schichtdicke,
- K_2 den Einfluss von Randeffekten und
- K_3 die Probengeometrie erfasst.

Bei der Widerstandsmessung an einer dünnen Schicht auf nicht leitendem Substrat oder auch bei Vorliegen von pn-Übergängen ergibt sich ein Korrekturfaktor K_1 . Er beruht darauf, dass durch den Übergang zwischen den Schichten der Probe eine gleichmäßige, kugelförmige Ausbreitung der Potentiallinien an den Spitzen 1 und 4 nicht mehr gegeben ist. Für sehr dünne Schichten mit Ausdehnungen geringer als ein halber Spitzenabstand gilt:

$$K_1 = \frac{d}{2s \cdot \ln(2)}. \quad (50)$$

Durch Einsetzen von K_1 in Gleichung (49) ergibt sich für den spezifischen Widerstand:

$$\rho = \frac{\pi d}{\ln(2)} \frac{U}{I}. \quad (51)$$

Mit Gleichung (43) ergibt sich damit der Flächenwiderstand im Falle einer dünnen Schicht auf nicht leitendem Substrat oder bei pn-Übergängen,

$$R_s = \frac{\pi}{\ln(2)} \frac{U}{I}. \quad (52)$$

Die weiteren Korrekturfaktoren betreffen die laterale Größe der endlich ausgedehnten Proben. Im Falle der Annäherung einer der Spitzen an den Rand des Wafers ändert sich das elektrische Feld zwischen den stromführenden Spitzen und damit auch die gemessene Potentialdifferenz. Unter Annahme von virtuellen Stromquellen lässt sich dieser Potentialunterschied berechnen. Dabei können zwei Fälle unterschieden werden,

- Die Spitzen sind parallel zu einem reflektierenden (nicht leitenden) geraden Rand.
- Die Spitzen sind senkrecht zu einem reflektierenden (nicht leitenden) geraden Rand.

Beim verwendeten Vierspitzenmessgerät sind die Spitzen senkrecht zum Rand ausgerichtet. Für den Korrekturfaktor ergibt sich

$$K_2 = \frac{1}{1 + \frac{1}{1 + 2 \cdot \frac{l}{s}} - \frac{1}{2 + 2 \cdot \frac{l}{s}} - \frac{1}{4 + 2 \cdot \frac{l}{s}} + \frac{1}{5 + 2 \cdot \frac{l}{s}}}. \quad (53)$$

Bei den gewöhnlichen Spitzenabständen von 1 bis 1,59 mm wird dieser Korrekturfaktor erst unterhalb von ca. 5 mm Randabstand signifikant kleiner als eins. Bei der Messung ist daher dieser Abstand auszuschließen oder der Korrekturfaktor zu berücksichtigen.

Betrachtet man nur relativ kleine Proben mit einem Durchmesser D , so kommt ein weiterer Korrekturfaktor hinzu, der sich wie folgt berechnet:

$$K_3 = \frac{\ln(2)}{\ln \left[2 \cdot \frac{(D/s)^2 + 3}{(D/s)^2 - 3} \right]}. \quad (54)$$

Für Messungen am ganzen Wafer wird dieser Korrekturfaktor jedoch nicht benötigt.

Duale Messkonfiguration

Für genaueste Vierspitzenmessungen können über die Verwendung von zwei Messkonfigurationen an derselben Messposition geometrische Effekte endlicher Probenabmessungen einschließlich der Randeffekte berücksichtigt werden. Solche Methode wird als Dual-Messung bezeichnet. Die erste Konfiguration ist die oben beschriebene mit dem eingprägten Stromfluss zwischen den Spitzen 1 und 4 und der Spannungsmessung an den Spitzen 2 und 3. In der zweiten Konfiguration wird der Stromfluss zwischen den Spitzen 1 und 3 eingestellt und die Spannung zwischen den Spitzen 2 und 4 bestimmt. Durch die zwei unabhängigen Messungen für eine Spitzenposition an der Oberfläche ergeben sich zusätzliche Informationen, die Messfehler einschränken. Es ergeben sich damit folgende Vorteile:

- Eine unbestimmte Ausrichtung der Spitzen bezüglich der Probengeometrie ist hierbei nicht nachteilig.
- Die Ausdehnung der Probe ist unerheblich, geometrische Einflüsse sind im Unterschied der Messungen enthalten.

- Die tatsächlichen Spitzenabstände untereinander fließen in die Messergebnisse ein. Abweichungen dieser Abstände gegenüber den nominellen werden so korrigiert.

Der Flächenwiderstand ergibt sich aus den Dual-Messungen wie folgt:

$$R_s = \left(-14,696 + 25,173 \left(\frac{R_a}{R_b} \right) - 7,872 \left(\frac{R_a}{R_b} \right)^2 \right) R_a, \quad (55)$$

mit

$$R_a = \frac{1}{2} \left(\frac{U_{v23}}{I_{v14}} + \frac{U_{r23}}{I_{r14}} \right) \quad \text{und} \quad R_b = \frac{1}{2} \left(\frac{U_{v24}}{I_{v13}} + \frac{U_{r24}}{I_{r13}} \right). \quad (56)$$

Dabei bezeichnen in den Indizes die Zahlen die Nummern der Spitzen, zwischen denen Strom und Spannung ermittelt werden, und v und r die Richtung des Stromflusses (vorwärts und rückwärts) bei den jeweiligen Messungen, über die bei dieser Methode gewöhnlich noch gemittelt wird.

Verwendetes Messsystem

Die Flächenwiderstandsmessungen an den dotierten Proben wurden an einem Vierspitzenmessgerät Map 168 von der Firma CDE durchgeführt. Für die Untersuchungen in der Reinraumumgebung standen die in der folgenden Tabelle angegebenen Messspitzen zur Verfügung.

Tabelle 4.8: Eigenschaften der vorhandenen Vierspitzen-Messköpfe.

Bezeichnung	Spitzenabstand	Spitzenradius	Aufpresskraft
C-Probe	1 mm	200 μm	0,98 N
E-Probe	1,58 mm	40 μm	1,96 N

Das Gerät besitzt ein automatisches Wafertransfersystem mit Waferentnahme aus einer Standardhorde. Vor der Ausführung von Messrezepten, in denen für die Bestimmung von Flächenwiderstandsverteilungen die Positionen der Einzelmessungen festgelegt werden können, erfolgt eine automatische Ausrichtung des Wafers für eine exakte Definition der Messpositionen bezüglich des Waferflats. Gewöhnlich wurden dann unter einer Abdeckung, die direkten Lichteinfall verhindert, an 81 oder 289 gleichmäßig über den Wafer verteilten Positionen die Einzelmessungen in Dual-Konfiguration ausgeführt.

Vor der Messaufgabe erfolgen automatisch eine Bestimmung des Probenkontakts und eine zusätzliche festgelegte Absenkung des Messkopfes (typisch 0,2 mm) für eine reproduzierbare Spitzenanpressung und eine Auswahl des passenden Referenzwiderstands für die Spannungsmessungen mit günstigster Verstärkung bei optimalem Stromfluss. Eine Einzelmessung besteht dabei aus der Auswertung des Anstiegs einer I - U -Kennlinie aus 2039 einzelnen Datenpaaren. Die maximalen Ströme und Spannungen liegen dabei typischerweise unter 0,7 mA und, je nach Widerstand (oder Dotierung), bei 0,2 bis unter 0,1 V. Die Güte der Einzelmessung wird dabei insbesondere aus der Linearität der I - U -Kennlinie durch einen so genannten Meritwert bestimmt. Dieser wird wie folgt ermittelt:

$$\text{Merit} = \frac{k \cdot N_{\text{Mess}} \cdot U_{\text{Range}} \cdot I_{\text{Range}} \cdot U_{\text{Range-Drive}}}{\chi^2} \quad (57)$$

Dabei sind N_{Mess} die Anzahl der gemessenen I - U -Paare, U_{Range} der Spannungsbereich zwischen Spitze 2 und 3, I_{Range} der Strombereich an Spitze 1 und 4, $U_{\text{Range-Drive}}$ der Bereich der für den Stromfluss erforderlichen Spannung und χ^2 die quadrierte mittlere Abweichung der Werte von der berechneten Regressionsgeraden. Die Konstante k wurde empirisch so ermittelt, dass ein Meritwert von eins für eine noch akzeptable Messung steht. Sehr gute Messungen ergeben Meritwerte von über 40. Durch die Auswertung dieses Wertes wird auch gewährleistet, dass ausschließlich ohmsches Verhalten des Widerstands zu einer gültigen Messung führt, verfälschende Einflüsse von nichtlinearen parasitären Effekten auf das Messergebnis werden damit ausgeschlossen.

Wiederholungsmessungen an einer Position mit und ohne Messkopfbewegung wurden durchgeführt, um die Reproduzierbarkeit der Einzelmessungen zu ermitteln. Bei 20 Messungen an aufgesetzt belassenen Spitzen wurde eine Standardabweichung für die Einzelmessung von unter 0,03 % ermittelt, bei 5 Messungen mit Wiederaufsetzen der Spitzen ist der Fehler kleiner als 0,3 % für gut konditionierte Spitzen. Hierbei ist zu beachten, dass mit wiederholtem Aufsetzen auf dieselbe Position eine starke Beanspruchung mit zunehmender Beschädigung der Oberfläche einhergeht [Clarysse '02].

4.4.2.2 Bestimmung des Dotierprofils durch Sekundärionen-Massenspektrometrie

Die Sekundärionen-Massenspektrometrie (SIMS) ist ein Verfahren zur tiefenaufgelösten Bestimmung der chemischen Konzentration von Spezies in einer festen Matrix. Es beruht im Wesentlichen auf dem Sputtern des Messobjektes mit einem Ionenstrahl, den Primärionen, und der darauf folgenden massenspektrometrischen Untersuchung der herausgelösten Sekundärionen. Damit können Tiefenprofile der chemischen Konzentration über mehrere Größenordnungen bestimmt werden. Für die Bestimmung von Dotierprofilen ist zu beachten, dass mit dem SIMS-Verfahren die gesamte chemische Konzentration des Dotierstoffs ermittelt wird, eine Unterscheidung zwischen elektrisch aktiven und inaktiven Dopanten ist damit nicht möglich.

Aufgrund der Bedeutung dieses Verfahrens für die Dotierprofilanalyse soll im Folgenden darauf noch etwas eingehender eingegangen werden, insbesondere im Hinblick auf die Phosphoranalyse in Silicium. Eine Einführung in die Grundlagen des SIMS-Verfahrens ist z.B. in [Wilson '89] zu finden. Zunächst sollen die in diesem Zusammenhang wichtigsten Begriffe kurz erläutert werden.

Ionisierungseffizienz:	Anteil <i>ionisierter</i> Teilchen in der Gesamtzahl der abgesputterten Teilchen,
Sekundärionenausbeute:	Anzahl von abgesputterten Sekundärionen einer Teilchensorte pro eintreffendem Primärion,
Sputterausbeute:	Gesamtzahl an abgesputterten Teilchen pro einfallendem Primärion,
Sputterrate:	Abtragsrate von Probenmaterial in nm/s,
Empfindlichkeit:	Abhängigkeit der jeweiligen Sekundärionenintensität von der im Substrat vorhandenen Atomdichte der entsprechenden Teilchensorte.

Für eine Analyse in Massenspektrometern müssen die Spezies ionisiert sein. Es werden daher solche Primärionen verwendet, die eine hohe Sekundärionenausbeute, das heißt eine möglichst große Anzahl von Sekundärionen pro einfallendem Primärion, erzielen.

Eine hohe Ausbeute an *positiven* Sekundärionen wird mit Sauerstoff-Primärionen (O_2^+) erreicht. Diese werden z.B. für die Analyse von Bor in Silicium verwendet, da die Ausbeute von positiven Borionen in der Siliciummatrix höher ist, als die der negativen. Bei der Phosphoranalyse in Silicium sind die Ausbeuten der negativen und positiven Phosphor-Sekundärionen ähnlich, beide Arten sind daher im Hinblick auf eine effektive Ausbeute für die Analyse geeignet.

Bei der Verwendung der *negativen* Sekundärionen werden in der Regel Cäsium-Primärionen (Cs^+) verwendet, da sie für diese die Ausbeute erhöhen. Für eine hohe Tiefenauflösung sind Primärionen mit einer geringeren Eindringtiefe besser geeignet. Schwerere Ionen besitzen eine geringere Eindringtiefe als leichte, so dass für die Phosphoranalyse oft die Cs^+ -Primärionen benutzt werden. Außerdem kann mit Cs^+ schneller gesputtert werden als mit O_2^+ .

Die Tiefenauflösung hängt auch von dem Auftreffwinkel der Primärionen und der Primärionenenergie ab. Diese Energie ist dabei die effektive Aufprallenergie, die sich aus der Spannung für die Beschleunigung und der Substratspannung für die Extraktion der Sekundärionen wie folgt ergibt:

$$E_{ip} = (U_{Quelle} - U_{Substrat}) q_{ip} . \quad (58)$$

Im Allgemeinen gilt, dass für größere Winkel bezüglich der Substratnormalen die Eindringtiefe abnimmt. Geringere Primärionenenergien haben auch geringere Eindringtiefen zur Folge, so dass damit höhere Auflösungen erreicht werden. Eine Beziehung für die Eindringtiefe in Abhängigkeit von der Primärionenenergie und dem Einfallswinkel wurde nach Extrapolation von publizierten Reichweitedaten für Cs^+ wie folgt angegeben [Vandervorst '87]:

$$z_E = 1,838 \text{ nm} \cdot \left(\frac{E_{ip}}{\text{keV}} \right)^{0,68} \cdot \cos \delta . \quad (59)$$

Die Eindringtiefe erhöht also sich mit Zunahme der Primärenergie E_{ip} und mit abnehmendem Einfallswinkel δ bezüglich der Substratoberflächennormalen. In diesem Zusammenhang ist der Einfluss der Primärionenenergie auf das Ionenstrahlmischen und auch die Aufrauung des Sputterkraters von Bedeutung, beide Effekte mindern die Tiefenauflösung [Wittmaack '84; Wittmaack '85]. Außerdem werden die Sekundärionenausbeute und die Sputterrate von der Primärionenenergie beeinflusst. Sie kann nun wiederum nicht beliebig gering gewählt werden, da eine höhere Energie einen höheren Primärstrom und eine bessere Fokussierung ermöglicht.

Bei der Auswertung der Messergebnisse sind Effekte zu beachten, die zu einer Abweichung des gemessenen Profils vom tatsächlichen führen. Um präzise Informationen der Konzentrationen in Abhängigkeit von der Tiefe im Substrat zu erhalten, sollen die Elemente natürlich an genau den Stellen analysiert werden, wo sie sich in der Matrix vor dem Messen befinden. Das heißt jedoch, dass jeder durch die Messung selbst verursachte Transportvorgang im Substrat das Ergebnis verfälscht und die Tiefenauflösung vermindert.

Begrenzung der Tiefenauflösung durch Ionenstrahlmischen (ion beam mixing)

Der Primärionenstrahl beeinflusst das zu analysierende Substrat, so dass durch die Messung selbst die Konzentrationsverläufe verändert werden können. Folgende, durch den Ionenstrahl hervorgerufene Transportvorgänge sind zu berücksichtigen:

- Mischen durch direkten Stoß (recoil mixing)

Durch direkten Stoß mit Primärionen werden Substratome in Strahlrichtung in größere Tiefen des Substrats transportiert. Als stark anisotroper Prozess beeinflusst er maßgeblich das Einmischen oberflächennaher Substratome in tiefere Schichten und vermindert somit die Tiefenauflösung. Die Tiefe des durchmischten Bereiches ist von der Eindringtiefe der Primärionen im Substrat abhängig [Vandervorst '87].

- Mischung in der Stoßkaskade (cascade mixing)

Eine Durchmischung tritt weiterhin durch sekundäre Stoßprozesse von bereits gestoßenen Substratomen mit noch ruhenden Atomen auf. Es bilden sich Stoßkaskaden, wobei nach wenigen Stoßprozessen die ursprüngliche Richtungsinformation verloren ist und isotropes Mischen stattfindet.

- Ionenbestrahlungsinduzierte Diffusion (radiation-enhanced diffusion)

Diffusionsprozesse sind stark abhängig von Defektkonzentrationen im Substrat. Durch den Ionenbeschuss bei der Messung werden Störstellen über die Gleichgewichtskonzentration hinaus im Substrat induziert, die die Diffusion der zu analysierenden Elemente selbst bei den Messtemperaturen signifikant ansteigen lassen können. Die Diffusionsrichtung ist neben dem Konzentrationsgradienten der diffundierenden Spezies stark abhängig von der Tiefenverteilung der Defekte.

Die gemessene Tiefenverteilung der Elemente im analysierten Material kann außer durch den Sputterprozess auch durch die Form des Sputterkraters von der tatsächlichen abweichen.

Insbesondere bei der Untersuchung von Substraten mit dünnen Schichten des zu analysierenden Materials an der Oberfläche kommt es durch das Ionenstrahlmischen zu Abweichungen der gemessenen Tiefenprofile vom tatsächlichen Verlauf. Für die folgende Beschreibung sei der Einfachheit halber angenommen, dass vor der Untersuchung das zu messende schichtbildende Material ausschließlich an der Oberfläche vorhanden ist. Zwei charakteristische Bereiche können dann anhand von unterschiedlich stark fallenden Konzentrationsverläufen unterschieden werden.

1. Ein anfänglicher starker Abfall geht einher mit der Entfernung der Schicht durch Absputtern oder auch Einmischen in oberflächennahe Substratbereiche.
2. Anschließend folgt ein charakteristischer geringerer exponentieller Abfall des gemessenen Konzentrationsverlaufes der analysierten Schichtsubstanz im Substrat.

Als Folge des Einmischens von Schichtmaterial ins Substrat durch die Messung selbst wird somit ein Anteil Oberflächenmaterial im Substrat gemessen, der ohne die Messung dort nicht vorhanden wäre. Dies muss natürlich bei der Auswertung der Profile in Betracht gezogen werden. Umgekehrt heißt dies aber auch, dass bei Auftreten dieser charakteristischen

Profilform auf eine ursprüngliche Oberflächenbelegung mit dem analysierten Element geschlossen werden kann.

Der exponentielle Abfall des Schichtelementsignals aufgrund der Einmischung kann durch eine charakteristische Länge beschrieben werden. In der Regel ist diese Abfall- oder Mischlänge λ definiert als der Tiefenabstand $z_2 - z_1$, bei dem die Intensität der Sekundärionen I (und damit der gemessenen Konzentration) um den Faktor $1/e$ gefallen ist, das heißt,

$$\frac{I_{z_1}}{I_{z_2}} = e^{-\frac{z_2 - z_1}{\lambda}}. \quad (60)$$

Die Mischlänge λ ist dabei im Bereich der Primärionenreichweite [Vandervorst '84]. Der langsame exponentielle Abfall zeigt damit an, dass das Oberflächenschichtmaterial über die Eindringtiefe der Primärionen mehr oder weniger homogen ins Substrat eingemischt wird. Für die Messung der Konzentration bei der aktuellen Tiefe wird hingegen nur ein flacher Schichtbereich an der momentanen Oberfläche abgesputtert. Dieser Bereich, aus dem die aktuellen Informationen stammen, entspricht der mittleren Tiefe, aus der die Sekundärionen das Substrat verlassen können (mean escape depth). Er erstreckt sich auf nur ungefähr 1-2 Monolagen, die nur einen geringen Anteil von ursprünglichem Schichtmaterial enthalten, während der Rest wiederum kontinuierlich durch das Sputtern tiefer ins Substrat gemischt wird.

Die Länge λ ist neben der Primärionenenergie auch abhängig von der Art des eingemischten zu analysierenden Materials [Wittmaack '84]. Im Allgemeinen kann von einem durch Ionenstrahlmischen verursachten exponentiellen Abfall der Sekundärionenintensität mit einer Länge λ innerhalb des 3 bis 4fachen der Primärionenreichweite ausgegangen werden.

Die Tiefenauflösung wird auch von der Morphologie der Oberfläche des analysierten Gebiets beeinflusst. Wenn die Sputterfläche nicht eben oder parallel zur Substratoberfläche ist, wird Material aus verschiedenen Tiefenlagen gleichzeitig gesputtert und analysiert, die Tiefeninformation ist damit nicht mehr eindeutig. Insbesondere durch den Sputterprozess selbst werden Oberflächenrauigkeiten geschaffen. Entsprechend führt bei fehlerhafter Einstellung des Analysebereiches das Sputtern der Kraterseitenwände zu Verfälschungen des Profils. Für ebene Sputterkrater sind gleichmäßige Primärionenströme und -stromdichten erforderlich.

Segregation

Durch den Sputterprozess können Transporteffekte bei Raumtemperatur zusätzlich zum physikalischen Mischen im Festkörper induziert werden, die als temperaturaktivierte Prozesse für die Mobilisierung der Spezies sonst nur bei hohen Temperaturen stattfinden, wie Diffusion und Segregation. Diese Effekte können das Ionenstrahlmischen deutlich verstärken. Die Transportvorgänge sind dabei so gerichtet, dass chemische Gradienten, wie sie an Oberflächen oder Grenzflächen auftreten können, minimiert werden, was in Konkurrenz zu den Konzentrationsgradienten stattfindet.

Insbesondere bei Vorhandensein von Oxidschichten oder bei Verwendung von O_2^+ -Primärionen tritt die Segregation auf. Das Bestreben einer An- oder Abreicherung an sauerstoffangereicherten Oberflächen kann mit dem Unterschied der Oxidationsenthalpie (Bildungsenthalpie ΔH_f des Oxids) von analysiertem Element und Matrixelement in Zusammenhang gebracht werden. So ist z.B. bei der Profilierung von Phosphor in Silicium die Bildung von SiO_2 ($\Delta H_f = -911,6$ kJ/mol) und die von P_2O_3 ($\Delta H_f = -820,6$ kJ/mol) bzw. von P_2O_5 ($\Delta H_f = -1493,0$ kJ/mol) zu vergleichen. Der Absolutwert der Bildungsenthalpie von SiO_2 ist zwar sehr

hoch, jedoch im Vergleich zur Bildung von P_2O_5 geringer, so dass sich hier eher das Matrixelement Silicium an der oxidierenden Oberfläche anreichert.

Bei isolierenden Schichten, die auch während der Oxidation beim O_2 -Sputtern gebildet werden, kann zusätzlich durch elektrische Aufladungen infolge des Ioneneintrags eine durch das elektrische Feld induzierte Diffusion auftreten, die zu Profilabweichungen führt.

Elementabhängiges Sputtern (preferential sputtering)

Wenn die Sputterraten der verschiedenen Elemente des Substrats sehr unterschiedlich sind, kann es im Sputterbereich zu einer Anreicherung derer kommen, die die geringsten Raten aufweisen. Dieser Effekt ist jedoch nur signifikant bei sehr großen Massendifferenzen der Elemente und bei Sputtern mit leichten Ionen geringer Energie. Bei gewöhnlichen SIMS-Bedingungen ist dieser Effekt zu vernachlässigen.

Matrix-Effekte

Wie oben bereits dargestellt, werden die Sekundärionen-Ausbeuten beim SIMS-Sputtern durch chemische Effekte gezielt erhöht. Elektronegative Elemente an der Sputteroberfläche können die Ausbeute von positiven Sekundärionen und elektropositive Elemente die Ausbeute von negativen Sekundärionen wesentlich erhöhen. So wird bei Siliciumsubstraten vorzugsweise mit O_2^+ - bzw. Cs^+ -Primärionen gearbeitet, die mit dem Sputtern in oberflächennahe Schichten implantiert werden.

Als Matrix-Effekt wird bezeichnet, dass Unterschiede in der Empfindlichkeit der Elemente auftreten in Abhängigkeit der Art der Matrix, in die sie eingebettet sind. Dies kann von Unterschieden der Ionisierungseffizienz oder der Sputterraten herrühren. Durch Kalibrierungen können die Empfindlichkeiten in bekannter Matrix bestimmt werden. Abweichungen von den Gleichgewichtswerten der Empfindlichkeit treten jedoch bei Sputterbeginn auf, wenn noch kein Gleichgewicht von implantierten zu wieder abgesputterten Primärteilchen besteht, oder auch beim Sputtern durch Grenzschichten, wenn sich die Sputterausbeute und damit auch die Gleichgewichtskonzentration der Primärteilchen im Substrat ändern kann, was wiederum eine empfindliche Änderung der Ionisierungseffizienz (derselben Sekundärionensorte) zur Folge haben kann.

Matrix-Effekte bei Vorhandensein von Sauerstoff besitzen eine besondere Bedeutung. Der Sauerstoff kann in Oxidschichten, so auch in natürlichen Oxiden, an der Oberfläche vorhanden sein oder durch die Messung hinzukommen, etwa bei der Verwendung von O_2^+ -Primärionen oder durch zusätzlichen Einlass von Sauerstoff durch ein definiertes Leck während der Messung. Eine immer vollständige Oxidation der Oberfläche während der Messung, das heißt eine Absättigung mit Sauerstoff, ist von Vorteil für die Stabilisierung und Reproduzierbarkeit der Sekundärionenraten.

Oberflächeneffekte

Eine besondere Rolle spielen Effekte an der ursprünglichen Oberfläche oder der oberflächennahen Schicht, da sie zu deutlichen Abweichungen vom tatsächlichen Profilverlauf führen können.

Eine natürliche Oxidschicht ist auf den meisten Proben vorhanden. Auf Silicium bildet sich an der frischen Oberfläche bei Kontakt mit Luft bereits bei Raumtemperatur zügig eine Oxidschicht von 1 bis 1,5 nm Dicke (siehe Kapitel 5.1.3). Aufgrund der Erhöhung der Ausbeute positiver Sekundärionen durch Sauerstoff erhält man bei der Auswertung dieser Ionen

in der Regel einen ausgeprägten Peak in deren Intensität an der Oberfläche (z.B. bei der Analyse von Bor in Silicium mit O_2^+ -Primärionen).

Außerdem stellt sich ein Gleichgewicht an Primärteilchen, die in das Substrat implantiert werden, erst bei einer Tiefe ein, die ihrer projektierten Reichweite entspricht. Bis zu dieser Tiefe muss erst abgesputtert werden, um sowohl die Ausbeute als auch die Sputterraten selbst zu stabilisieren. So sollte das Sauerstoffprofil bei solchen Proben mit gemessen werden, um die Oxidbedeckung zu ermitteln. Günstig erweist sich hier auch eine Sauerstoffzufuhr über ein Leck oder das vorherige Aufbringen einer Schutzschicht mit einer Dicke größer als die Ionenreichweite zur Verhinderung der Bildung eines natürlichen Oxids.

Bei der Analyse der *negativen* Sekundärionen bei Verwendung von Cs^+ ist ein Oberflächenpeak durch Ausbeuteerhöhung aufgrund einer natürlichen Oxidschicht nicht ausgeprägt. Jedoch tritt auch hier ein Übergangsbereich zu Beginn der Messung an der ursprünglichen Oberfläche auf, in dem die Cs^+ -Konzentration, die für die Ausbeuteerhöhung erforderlich ist, noch nicht im Gleichgewicht ist. Diese „Gleichgewichtstiefe“ hängt auch hier von der Reichweite der Cs^+ -Primärionen und damit von ihrer Energie ab. Bei der Verwendung von Ionen der Energie von 2 bis 7 keV steigt diese Tiefe von 1 auf 4 nm an.

Messbedingungen für die Untersuchungen

Zur Erstellung hochaufgelöster Tiefenprofile der Phosphorkonzentration in Silicium mussten die Wafer zerteilt und aus dem Reinraum ausgeschleust werden. Um einen direkten Vergleich der Messergebnisse aus den verschiedenen Analyseverfahren zu ermöglichen, wurden vor dem Brechen der Wafer die ortsaufgelösten Messungen von Flächenwiderstand und Schichtdicke durchgeführt. Für die SIMS-Messungen wurden dann Proben von festgelegten Waferstellen ausgewählt.

Ein Großteil der Messungen wurde bei der Firma RTG Mikroanalyse GmbH Berlin durchgeführt. Dafür wurde ein IMS 4F-E6 (Fa. CAMECA) benutzt. Für erste Messungen wurden Cs^+ -Ionen mit einer effektiven Energie von 14,5 keV verwendet (10 kV Beschleunigungsspannung plus 4,5 kV Extraktionsspannung für die negativen Sekundärionen). Dabei zeigte sich, dass diese Primärionenenergie zu hoch für die geforderte Tiefenauflösung war. Daraufhin wurden die weiteren Messungen mit der geringstmöglichen Primärionenenergie von 3,5 keV durchgeführt. Für die Phosphorbestimmung in Silicium musste mit einer hohen Massenauflösung gearbeitet werden, um eine Trennung von $^{31}P^-$ (30,9738 amu) und $^{30}Si^1H^-$ (30,9816 amu) zu erreichen.

Zusätzliche Messungen an einigen Proben konnten mit einer ATOMIKA 6500 (Fa. PERKIN ELMAR) mit Cs^+ -Primärionen einer Energie von 4,5 keV für eine gleichzeitige Profilierung von Phosphor, Sauerstoff und Silicium am Institut für Physikalische Elektronik der Universität Stuttgart durchgeführt werden. Weitere Messungen bei einer geringen Primärionenenergie von 0,5 keV wurden am IHP in Frankfurt (Oder) durchgeführt.

5. Ergebnisse und Diskussion

5.1 Vorbehandlung der Siliciumwafer

Vorbereitende Experimente dienten der Charakterisierung der verwendeten Reinigungsschritte und der Oxidentfernung im HF-Dip. Dazu wurde die Oberfläche nach den jeweiligen Behandlungsschritten mittels Ellipsometer untersucht. Das Ziel dieser Untersuchungen war eine möglichst genaue Kenntnis der Siliciumoberfläche vor der Dotierung, um die Beherrschbarkeit von reproduzierbaren Vorbehandlungen für reproduzierbare Dotierergebnisse abzuschätzen, um die Dotierergebnisse in Bezug auf die anfängliche Oberfläche einzuschätzen und um einen Vergleich der Oberflächeneigenschaften vor und nach der Dotierung zu ermöglichen, da in-situ Messungen von Oberflächeneigenschaften während der Dotierung in der zur Verfügung stehenden RTP-Anlage nicht durchgeführt werden konnten.

So wurde untersucht, ob Oxidschichten auf den Siliciumwafern vor der Dotierung vorlagen, welche Dicken diese aufwiesen und mit welcher Geschwindigkeit eine Oxidation erfolgte. Außerdem wurden einige Vergleiche zwischen verschiedenen Reinigungsverfahren und HF-Dips angestellt.

5.1.1 Einschätzung der ellipsometrischen Messungen

Zur Charakterisierung der Oberfläche wurden Messungen am Ellipsometer durchgeführt, da diese eine zerstörungsfreie, empfindliche, relativ einfache und schnelle Charakterisierung der optischen Eigenschaften der Oberfläche erlaubte. Aus diesen optischen Eigenschaften sollten Rückschlüsse auf Veränderungen der Oberfläche durch Belegungen während der Vorbehandlung der Wafer erfolgen und die vor der Dotierung tatsächlich vorliegenden Oberflächeneigenschaften bewertet werden.

Dazu war jedoch eine kritische Einschätzung der Möglichkeiten und Beschränkungen der Messungen erforderlich. Das verwendete Ellipsometer besaß einen Messbereich für Wellenlängen von 250-870 nm. So sollte damit auch der Bereich von 250-380 nm zugänglich sein, im dem nach Literaturangaben für dünnste Schichten auf Silicium die größte Empfindlichkeit zu erwarten war [Kim '95]. Es zeigte sich jedoch, dass die Messwerte in diesem Bereich deutlich von den zu erwartenden abwichen. Ein Beispiel für gemessene Spektren von Ψ und Δ für den gesamten möglichen Messbereich unmittelbar nach einem HF-Dip eines vorher gereinigten Wafers ist in Abbildung 5.1 angegeben.

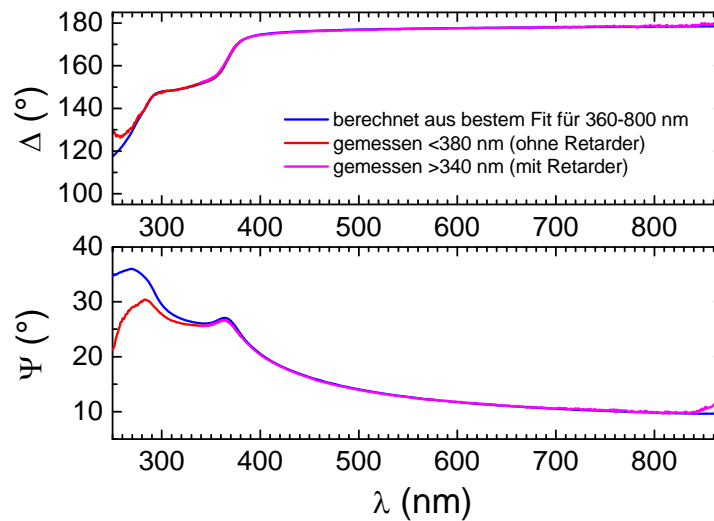


Abbildung 5.1: Gemessene Spektren von Ψ und Δ für den gesamten zugänglichen Messbereich von 250-870 nm unmittelbar nach HF-Dip eines mit Caro'scher Säure gereinigten Siliciumwafers im Vergleich mit zu erwartenden Verläufen, ermittelt aus bestem Fit für den Bereich von 340-800 nm.

In dieser Abbildung sind zwei Messbereiche überlagert dargestellt. Der für genaue Δ -Messungen nahe 180° zur Verfügung stehende Retarder ist erst für Wellenlängen oberhalb von etwa 340 nm einsetzbar, Licht geringerer Wellenlängen wird hier zu stark absorbiert. So ist der Bereich von 340-870 nm mit Retarder und nur der Bereich von 250-380 nm in einer separaten Messung ohne Retarder aufgenommen worden. Es zeigte sich, dass die Aufnahme des gesamten Messbereichs ohne den Retarder zu großen Fehlern der Δ -Messwerte nahe 180° führte. In der Abbildung ist zu erkennen, dass die Werte beider Messungen im Überlagerungsbereich von 340-380 nm gut übereinstimmen.

Es wird nun in dieser Abbildung deutlich, dass die Messwerte unterhalb von ca. 340 nm so stark fehlerbehaftet sein müssen, dass sie für weitere Auswertungen nicht verwendet werden können. Bei Verwendung des einfachen Modells einer Oxidschicht auf Silicium (siehe Abbildung 4.9) ist es sehr gut möglich, den Bereich von 360-800 nm sinnvoll zu approximieren, es ergibt sich dann hierfür eine Oxidschichtdicke von 0,61 nm und der blau dargestellte Verlauf der theoretischen Werte für Ψ und Δ für dieses Modell. Die Fehler im Bereich der kürzeren Wellenlängen sind somit deutlich erkennbar, der hier gemessene Verlauf kann nicht allein über ein sinnvolles Modell von optisch aktiven Oberflächenschichten interpretiert werden. Es zeigte sich, dass in diesem Bereich auch nur geringe Intensitäten gemessen werden konnten, was auf zu große Absorption in den Lichtleitkabeln hindeutete. Somit konnte der Messbereich von 250 bis 340 nm nicht für die weiteren Auswertungen verwendet werden. Aufgrund von erhöhten Schwankungen der Messwerte oberhalb von ca. 800 nm wurde der Messbereich zu höheren Wellenlängen hin auf 800 nm begrenzt, was an dieser Stelle keine Einschränkung bedeutet, da hier kaum zusätzliche Informationen enthalten sind.

Anhand von wiederholten Messungen, ohne die Probe zu verändern, konnte festgestellt werden, dass in dem gewählten Wellenlängenbereich die ellipsometrischen Messungen äußerst reproduzierbar durchgeführt werden konnten, die so ermittelten Spektren zeigten keine signifikanten Unterschiede.

5.1.2 Vergleich der Reinigungsschritte

Mit Testwafern wurden verschiedene Vorbehandlungsschritte durchgeführt und die Oberflächen unmittelbar nach den einzelnen Schritten mittels Ellipsometer untersucht. Dafür wurde jeweils in der Mitte der Wafer im Wellenlängenbereich von 340-800 nm gemessen. Bei den Untersuchungen nach den HF-Behandlungen wurde darauf geachtet, dass die Messungen möglichst unmittelbar nach dem HF-Schritt, das heißt nach dem Trocknen der Wafer, durchgeführt wurden. Für die Entnahme aus der Trockenschleuder und Justierung des Wafers unter dem Ellipsometer sind ca. 3 Minuten bis zur ersten Messung erforderlich.

Das optische System der Oberfläche wurde wie in Abbildung 4.9 dargestellt modelliert und darin die Schichtdicke der Oxidschicht als einziger Parameter variiert, bis die beste Übereinstimmung zwischen den gemessenen und berechneten Werten gefunden wurde. Ein Beispiel für eine Messung, die direkt nach der Oxidentfernung durch HF-Dip aufgenommen wurde, zeigt die Abbildung 5.2.

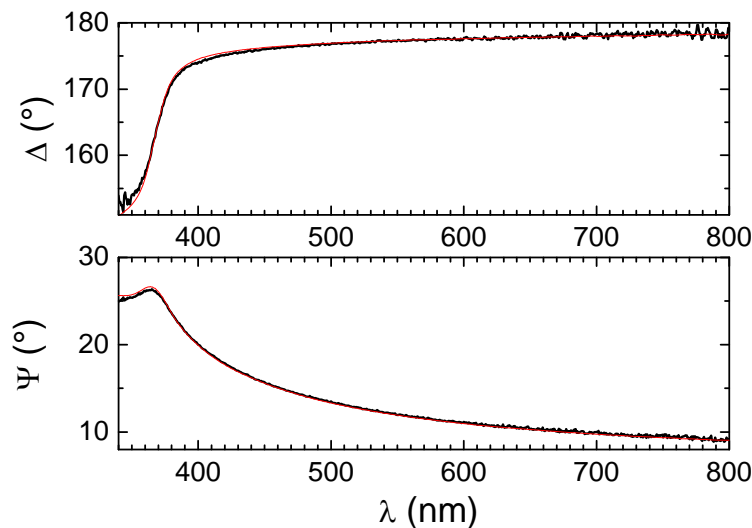


Abbildung 5.2: Spektren von Ψ und Δ für Wafer A-1 unmittelbar nach einem HF-Dip und berechnete Werte für besten Fit mit 0,57 nm Oxiddicke.

Allgemein konnte festgestellt werden, dass bei allen Messungen nach den Vorbehandlungen Spektren wie in Abbildung 5.2 dargestellt erhalten wurden, die sich gut mit dem einfachen Modell beschreiben ließen. Die daraus ermittelten Werte der (äquivalenten) Oxiddicken nach den einzelnen Vorbehandlungsschritten sind in der Tabelle 5.1 zusammengestellt.

Tabelle 5.1: Zusammenfassung der ermittelten (äquivalenten) Oxiddicken nach verschiedenen Vorbehandlungsschritten an Siliciumwafern.

Wafer-Nr.	Reinigungsschritt	HF-Dip	d_{Oxid} in nm			Bemerkung
			unbehandelt	nach Reinigung	nach HF	
A-1	Caro 120 °C, 5 min	60 s, 1 %	0,8	1,46	0,57	
A-2	Caro 120 °C, 5 min	120 s, 1 %		1,17	0,53	
A-3	Caro 120 °C, 5 min	30 s, 1 %		1,22	0,57	
B-1	SC1 80 °C, 5 min	60 s, 1 %		1,11	0,66	ohne Spülen
B-2	SC1 80 °C, 5 min	60 s, 1 %	1,09	1,17	0,55	
C-1	Caro 120 °C, 5 min	300 s, 0,1 %	1,66	1,40	0,58	
C-2	Caro 120 °C, 5 min	600 s, 0,1 %	1,77	1,50	0,64	

Aus diesen Ergebnissen lässt sich folgendes ableiten.

Oxidentfernung

Durch einen HF-Dip konnte in keinem Fall eine geringere (äquivalente) Schichtdicke als ca. 0,5 nm erreicht werden. Insbesondere konnte weder durch längeres Ätzen in HF eine deutliche Reduzierung dieses Wertes erzielt werden, noch konnte durch Auslassen des Spülens nach dem HF eine Verminderung erreicht werden. Vielmehr zeigte sich, dass ohne das Spülen sogar eine größere (äquivalente) Schichtdicke gemessen wurde. Dabei wurde der Wafer direkt nach Entnahme aus dem HF-Bad, wonach sich eine hydrophobe Siliciumoberfläche ausbildet, so dass der Wafer praktisch filmfrei und trocken ist, vermessen. Ein doppelt so langer HF-Dip (Wafer A-2 gegenüber Wafer A-1) ergab keine signifikante Verminderung der Oxiddicke. Und auch ein längeres Belassen eines Wafers im HF-Bad (selber Wafer A-1 nach einem folgenden HF-Dip) führte zu keiner bedeutenden Änderung der Eigenschaften. Nach 18 Stunden im HF wurden 0,73 nm ermittelt, der gleiche Wert, wie nach einem vorherigen wiederholten Dip von nur einer Minute. Nach mehrmaliger Wiederholung der HF-Behandlung ohne weitere Reinigung war die ermittelte Schichtdicke jedoch immer größer als nach dem ersten Dip.

Aus den Ergebnissen kann geschlossen werden, dass

- es entweder nicht möglich war, die Oxidschicht vollständig zu entfernen,
- eine frische Oxidschicht unmittelbar nach dem HF-Dip oder noch während des HF-Dips wächst
- oder aber der ermittelte Wert keine Oxidschicht repräsentiert, sondern tatsächlich nur eine „äquivalente“ Oxidschicht (siehe Kapitel 5.1.4).

Ohne nach dem HF-Dip folgendes Spülen zeigte sich eine größere ermittelte Schichtdicke, was auf eine Kontamination mit noch an der Oberfläche haftenden Fluorverbindungen hindeuten kann, die als Reaktionsprodukte bei der Wasserstoff-Terminierung der Oberfläche entstehen, wie sie auch Gräf et al. in Untersuchungen nachweisen konnte [Gräf '89]. Ein Indiz dafür ist auch, dass der Widerstand des Spülwassers bei Beginn des Spülens deutlich ansteigt. Ein Spülen nach dem HF-Dip ist somit günstiger.

Andererseits ist die Wasserstoff-Terminierung der Oberfläche kein effektiver Schutz vor Oxidation des Siliciums [Angermann '04]. Vielmehr ist anzunehmen, dass eine Oxidation, die an den rückwärtigen Bindungen der terminierten Oberflächen-Siliciumatome einsetzt, bereits im Spülwasser erfolgt. Dies könnte dann zu der vor Beginn der Messungen vorhandenen Oxiddicke führen. So konnte auch gefunden werden, dass bei Belassen des Wafers im DI-Wasser-Spülbad eine deutlich schnellere Oxidation erfolgte als in der gleichen Zeit an Luft. Ein direkter Vergleich von Wafer A-1 und A-2 nach nochmaligem HF-Dip, wobei Wafer A-1 nach dem Spülen an Luft vermessen wurde, während der andere Wafer A-2 im Wasser verblieb, ergab nach 17,5 Stunden an Luft eine Oxiddicke von 1,14 nm und nach der gleichen Zeit in Wasser eine Oxiddicke von 1,42 nm. (Auch zeigte derselbe Wafer A-2 nach nochmaligem HF-Dip und Lagerung von 17,5 Stunden, diesmal an Luft, nur eine Oxiddicke von 1,22 nm.) Offensichtlich ist das verwendete DI-Wasser zwar für die Entfernung der Kontaminationen aus dem HF-Bad und die H-Terminierung der Oberfläche von Vorteil, jedoch kein Mittel, die Oxidation zu verzögern. Es ist bekannt, dass der Sauerstoffgehalt des Wassers von großer Bedeutung für die Oxidation ist. Je geringer der Gehalt, desto langsamer sollte auch die Oxidation ablaufen. Wie von Watanabe und Sugita beschrieben, ist die Sauerstoffkonzentration sogar für eine Umkehrung der Oxidation in einen Ätzprozess mitbestimmend, ebenso wie die Temperatur des Bades [Watanabe '95]. Bei dem verwendeten DI-Wasser konnte der Sauerstoffgehalt weder bestimmt noch reduziert werden.

Rauigkeit

Auf eindeutige, signifikante Veränderungen der Rauigkeit der Oberfläche in Abhängigkeit von den HF-Behandlungen konnte aus den Messungen nicht geschlossen werden. Insbesondere die Behandlungen in HF-Bädern der Konzentration von 0,1 und 1% sind dieser Hinsicht nicht kritisch für die Oberfläche, wie der Versuch mit der 18-stündigen HF-Behandlung zeigte. Jedoch konnte festgestellt werden, dass durch natürliche Oxidation nach HF-Oxidentfernung die Oberflächeneigenschaften geringfügig verändert wurden, da nach wiederholter Entfernung dieser Oxidschicht der ursprüngliche geringste Wert der Oxiddicke nicht mehr erreicht werden konnte. Dies war insbesondere nach der Oxidation im DI-Wasserbad der Fall.

Reproduzierbarkeit

Wie in obiger Zusammenfassung ersichtlich, wurde im allgemeinen nach den HF-Dips eine gute Reproduzierbarkeit der Oberflächeneigenschaften festgestellt, wobei auch die vorher durchgeführten Reinigungsschritte keinen wesentlichen Einfluss darauf zu haben scheinen, was an den ähnlichen Oxiddicken nach der chemischen Oxidation im Reinigungsbad ablesbar ist.

Bei den Untersuchungen zeigte sich jedoch, dass ein bedeutender Einfluss auf die Oberflächeneigenschaften durch die weitere Oxidation der H-terminierten Oberfläche zu verzeichnen ist. Die Stabilität der Terminierung gegenüber Oxidation wurde daraufhin in zeitabhängigen ellipsometrischen Messungen untersucht.

5.1.3 Stabilität der Oberfläche und Oxidwachstum an Luft

Da ein Kontakt der Wafer mit der Reinraumluft nach der Reinigung unmittelbar vor dem Beladen in die Vakuumschleuse prinzipiell nicht zu vermeiden war, sollte abgeschätzt werden, in welcher Zeit das weitere Wachstum der Oxidschicht voranschreitet. Dazu wurden mehrere Versuche durchgeführt, in denen Siliciumwafer unmittelbar nach der Reinigung mittels Ellipsometer zeitabhängig vermessen wurden. Startwert für die Zeitdauer der Luft-einwirkung war dabei der Zeitpunkt, an dem die Wafer aus der mit Stickstoff gespülten Trockenschleuder entnommen wurden. Nach Justage des Wafers unter dem Ellipsometer wurde automatisch nach vorher in einem Messprogramm festgelegten Zeitabschnitten ein komplettes Ψ - Δ -Spektrum im Wellenlängenbereich von 340-800 nm aufgenommen und mittels des einfachen Modells nach Abbildung 4.9 die Schichtdicke des wachsenden Siliciumoxids ermittelt und aufgezeichnet.

Der Anstieg des Wertes der Oxiddicke war damit sehr genau nachzuweisen. Wie oben bereits ausgeführt, sollte dabei die relative Änderung des Schichtdickenwertes genauer zugänglich sein, als der jeweilige Absolutwert selbst. In der Abbildung 5.3 ist als Beispiel für solche Messungen der Anstieg der Oxiddicke von Wafer A-3 angegeben.

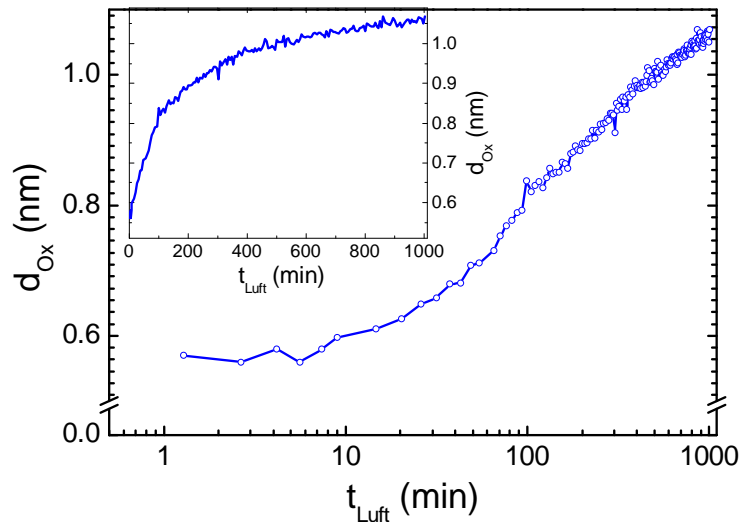


Abbildung 5.3: Anstieg der Schichtdicke von Wafer A-3 an Luft nach einer HF-Behandlung, Darstellung im Logarithmus der Zeit an Luft. Die kleine Darstellung zeigt den Verlauf mit linearer Zeitachse).

Für die Hauptdarstellung in der Abbildung ist die Zeitachse logarithmisch eingeteilt, und zum Vergleich ist in der kleinen Darstellung der Verlauf mit linearer Zeitachse gezeigt. Hier ist gut zu erkennen, dass nach einer gewissen „Verzögerungszeit“ von nur ca. 10 min, in der die Schichtdicke sich noch kaum verändert, ein deutlicher Anstieg der Oxiddicke einsetzt, der ab einer Zeit von ungefähr einer Stunde einem logarithmischen Wachstum mit der Zeit entspricht.

Offenbar ist die Wasserstoff-Terminierung der Oberfläche oder eine die geringe Oxidbedeckung nach der Reinigung kein Schutz vor weiterer Oxidation. Die Zeit zwischen dem Beenden der Waferreinigung, d.h. der Entnahme der Wafer aus der Trockenschleuder, und dem anschließenden Beladen in die Schleuse sollte daher so kurz wie möglich sein, so dass für die Wafer gleiche Anfangsbedingungen für die jeweiligen Dotierprozesse vorliegen. Für die nachfolgenden Dotierversuche wurden Beladezeiten von unter 5 min bis zum Evakuieren der Schleuse eingehalten.

5.1.4 Analyse der Spektren

Zusätzlich wurde versucht, mittels einer eingehenderen Analyse der gemessenen Spektren die Qualität der Messungen und auch die von verschiedenen Modellen zu prüfen, um so mehr Informationen über Veränderungen oder den tatsächlichen Aufbau der Oberfläche in Abhängigkeit von den verschiedenen Behandlungsschritten zu gewinnen. Auffällig war bei genauerem Vergleich von gemessenen und modellierten Spektren, dass bei der Verwendung des einfachen Modells nach Abbildung 4.9 in beiden Spektren noch merkliche Abweichungen im Bereich von Wellenlängen kleiner etwa 380 nm und auch auffälligerweise im Δ -Spektrum im Bereich um die 400 nm auftraten (siehe Abbildung 5.2). Durch Einbeziehung von verschiedenen Interface-Schichten in die Modelle wurde nun versucht, diese Abweichungen zu minimieren [Herzinger '98]. Als Maß für diese Abweichungen wurde der mittlere quadratische Fehler (MSE) ausgewertet. Nach Prüfung von zahlreichen Modellen wurden zwei herausgefunden, mit denen sich unter der Annahme von auch zu erwartenden wahrscheinlichen Oberflächenverhältnissen eine Verringerung der Abweichungen erzielen ließ und die hier dem einfachen Modell gegenübergestellt werden sollen.

Dies sind

- die Annahme einer Zwischenschicht zwischen SiO_2 und Si als Bereich einer Einmischung von SiO_2 in Si, modelliert als Brüggemann-Modell mit einem Anteil f von SiO_2 in Si und der Interface-Dicke d_{Interf} als zusätzliche variierbare Parameter (siehe [Herzinger '98]) und
- die Annahme einer Zwischenschicht zwischen SiO_2 und Si aus amorphem Silicium, mit nur der Interface-Dicke $d_{a\text{-Si}}$ als zusätzlichem variierbaren Parameter (siehe z.B. in [Hu '97; Suni '98]).

Hierbei wurde durchgehend das oben beschriebene Cauchy-Modell für eine Oxidschicht verwendet. Andere Oxidmodelle lieferten qualitativ die gleichen Resultate. Eine Anpassung des Oxidmodells selbst, über eine Variation der Cauchy-Parameter (oder auch entsprechender Parameter bei der Verwendung eines Sellmeyer-Modells), erbrachten bei den offenbar vorliegenden dünnsten Oberflächenschichten, keine sinnvollen Ergebnisse¹. Bei dem Modell für amorphes Silicium handelte es sich um tabellarische Ψ - Δ -Werte nach Aspnes aus der Modelldatenbank der SpectraRay-Software.

Unter Einbeziehung des gesamten gewählten Messbereiches von 340 nm bis 800 nm war es mit keinem gewählten Modell gelungen, die Abweichungen im Bereich der Wellenlängen kleiner als etwa 380 nm deutlich zu verringern. Diese Abweichungen sind offenbar noch auf die oben erwähnten Messfehler des Ellipsometers in Bereich kurzer Wellenlängen zurückzuführen.

In der Abbildung 5.4 ist nun der Bereich von 400-600 nm dargestellt, von dem angenommen wurde, dass er durchgehend am genauesten gemessen werden konnte, und in dem die Abweichungen der Messwerte gegenüber dem einfachen Modell am deutlichsten waren. Dies betraf alle Messungen nach den verschiedenen Behandlungsschritten mehr oder weniger deutlich ausgeprägt.

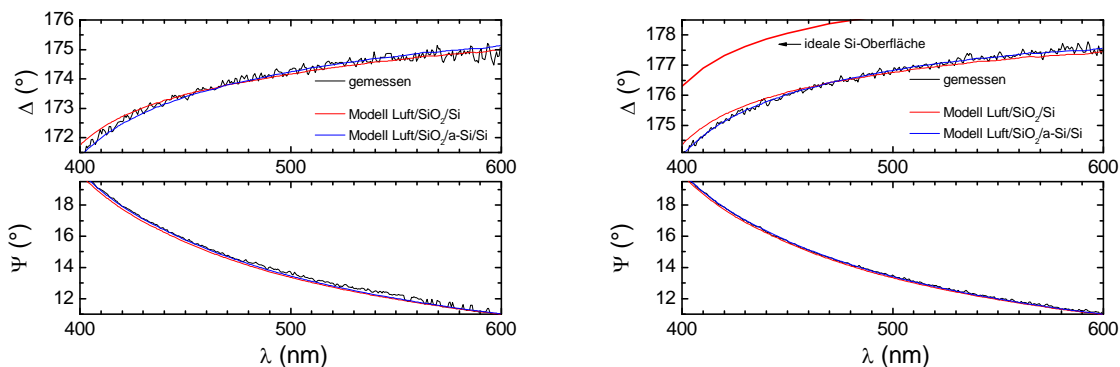


Abbildung 5.4: Ellipsometrische Messwerte im Bereich von 400-600 nm an Si-Waferoberflächen nach Caro-Reinigung (links) und unmittelbar nach HF-Behandlung (rechts) zusammen mit den theoretischen Verläufen aus der besten Anpassung des Modells mit einer Oxidschicht (rot) und mit zusätzlichem Interface aus amorphem Silicium (blau). Rechts mit dargestellt ist Verlauf von Δ bei idealer, reiner Si-Oberfläche.

In der linken Darstellung von Abbildung 5.4 sind die Spektren nach der Caro-Reinigung dargestellt und in der rechten die, die unmittelbar nach HF-Dip, Spülen und Trocknen aufgenommen wurden. Nach dieser Behandlung war die geringstmögliche Oxidbedeckung erreicht

¹ Wie eingangs in den Vorbetrachtungen zu den Beschränkungen der Ellipsometrie für dünne Schichten gezeigt, ist dies nicht verwunderlich. So kann bei genügender Anzahl veränderlicher Parameter jeder gemessene Verlauf „angepasst“ werden, ohne dass das Modell dann noch die tatsächliche Struktur zu repräsentieren.

worden. In dieser Abbildung ist zum Vergleich der Größe der Abweichungen und auch zur Verdeutlichung des Einflusses selbst so geringer Bedeckungen, wie die oben ermittelten, auf die ermittelten Spektren auch die Lage der Δ -Werte angegeben, wie sie für die reine, unbedeckte Siliciumoberfläche zu erwarten wären. Für die Ψ -Werte wäre selbst in dieser Darstellung keine Abweichung im Vergleich zur reinen Oberfläche deutlich.

Aus den Darstellungen ist zu ersehen, dass zwar bereits des einfache Modell eine gute Annäherung ermöglicht, dass jedoch allein mit der (plausiblen) Annahme einer Zwischenschicht, das erweiterte Modell eine fast exakte Abbildung der gemessenen Verlaufs der Δ -Werte liefert und selbst für die Ψ -Werte eine größere Übereinstimmung auftritt. In den beiden Abbildungen sind die Verläufe für das Modell mit dem Brüggemann-Interface nicht mit dargestellt. Sie ergaben zwar auch eine bessere Anpassung als das einfache Modell, jedoch nicht die Qualität der blau dargestellten. Darüber hinaus konnte festgestellt werden, dass für dieses Modell bei der Annahme einer Einmischung von Luft anstelle von Oxid ins Silicium, wie es bei einer rauen Oberfläche anzunehmen wäre, beide Modelle ähnliche Ergebnisse liefern, jedoch das Modell mit dem Oxid immer einen etwas geringeren MSE-Wert ergibt. Das gleiche wurde auch bei Annahme von polykristallinem Silicium anstelle von amorphem in dem anderen Modell festgestellt.

Anzumerken ist hier, dass sich diese Verbesserungen der Anpassungen nicht durch Variation von Parameter des Messgerätes selbst ergeben haben. Die Variation eines Winkel- oder Wellenlängen-Offsets des Ellipsometers, die mit der Software möglich ist und für Kalibrierzwecke genutzt wird, ergibt hier keine Verbesserungen, d.h. solche Offsets können an dieser Stelle ausgeschlossen werden.

Der Vergleich der MSE-Werte, als Maß für die Übereinstimmung der Spektren, zusammen mit den aus der Variation gewonnenen besten Werte der freien Parameter sind für die Messungen am Wafer A-1 nach den verschiedenen Behandlungsschritten in der Tabelle 5.2 angegeben.

Tabelle 5.2: Vergleich dreier Schichtmodelle für die Auswertung ellipsometrischer Messungen im Wellenlängenbereich von 400-600 nm nach verschiedenen Oberflächenbehandlungen von Silicium (BG: Abkürzung für Brüggemann).

Messung (erwartete Oberfläche)	Luft/SiO ₂ /Si		Luft/SiO ₂ /a-Si/Si			Luft/SiO ₂ /BG-Interface/Si			
	d_{Oxid} (nm)	MSE	d_{Oxid} (nm)	d_{a-Si} (nm)	MSE	d_{Oxid} (nm)	$d_{Interf.}$ (nm)	$f_{Interface}$	MSE
Eingangsmessung (Oxid)	0,87	0,0892	0,94	0,14	0,0400	0	0,7	0,653	0,0705
nach Caro-Reinigung (chemisches Oxid)	1,43	0,1133	1,50	0,15	0,0724	0	1,15	0,780	0,1011
3 min nach HF-Dip (kein Oxid)	0,61	0,0890	0,67	0,13	0,0447	0	0,5	0,627	0,0738
28 h nach HF-Dip (natürliches Oxid)	0,94	0,0841	1,00	0,12	0,0460	0	0,75	0,698	0,0705
nach 18 h in HF (raue OF)	0,78	0,0646	0,81	0,07	0,0487	0	0,62	0,678	0,0530

Aus der Tabelle ist zu ersehen, dass das Modell mit einer amorphen Zwischenschicht durchgehend die deutlich geringsten MSE-Werte und damit die beste Anpassung liefert. Die damit ermittelten Werte für die Oxidschicht weichen jedoch nicht sehr von denen ab, die mit dem einfachen Modell ermittelt wurden. So wäre bei Vorhandensein einer Interfaceschicht von 0,12 – 0,15 nm Ausdehnung im Modell mit a-Si-Interface die Oxidschicht nur 0,06 - 0,07 nm

dicker. Die Werte für die Oxidbedeckung nach HF-Dip wären jeweils mit ca. 0,7 nm im Bereich von 2 Monolagen SiO_2 [Kaneta '99], was somit die geringste mit den Vorbehandlungsschritten erreichbare Oxidbedeckung wäre. Die Ausdehnung des Interfaces wäre dabei im Bereich von Oberflächenstufen von 0,13 nm, wie sie bei Watanabe für die Oxidation gefunden worden sind [Watanabe '98]. Interessant ist, dass dies Interface nach den verschiedenen Behandlungen in etwa gleichermaßen gefunden wurde, jedoch nach der 18-stündigen HF-Behandlung das einfache SiO_2 -Modell bereits eine sehr gute Anpassung liefert und daher auch die ermittelte Interfacedicke hier mit 0,07 nm den geringsten Wert aufweist. Eine deutliche Veränderung der Oberflächeneigenschaften ließe sich hieran jedoch nicht erkennen. Auch scheint durch die natürliche Oxidation von 28 h in der Reinraumluft außer der wachsenden Oxidschicht keine weitere signifikante Veränderung der Oberfläche aufzutreten.

Unter der Annahme des Modells mit dem Brüggemann-Interface ergeben sich zwar auch geringere MSE-Werte als für das einfache Modell, jedoch nicht so geringe, wie das mit der amorphen Schicht. Interessant ist, dass bei dieser Modellannahme die beste Anpassung kein eigentliches Interface ergibt. Die Schichtdicke der angenommenen durchgehenden Oxidschicht ist jeweils Null, und das Modell besteht aus nur einer Schicht auf dem Silicium, die eine Mischung von Oxid und Silicium darstellt. Dies könnte man als „effektive“ Oxidschicht betrachten, mit jeweils verschiedener Dicke aber auch unterschiedlichem Anteil von oxidiertem Silicium. So sind hier z.B. nach Caro-Reinigung und auch nach 28-stündiger Oxidation an Luft nicht nur die Schichtdicken dieser Schicht größer, sondern auch der Anteil von oxidiertem Silicium. Jedoch lassen sich auch unter Verwendung dieses Modells keine grundlegenden Veränderungen der Struktur der Oberfläche, wie eine veränderte Rauigkeit, erkennen.

Bei Vergleich der Werte für die Schichtdicke aus Tabelle 5.1 für den Wafer A-1, die aus dem Fit unter Verwendung des einfachen Modells über den gesamten gewählten Messbereich gewonnen wurden, mit den ermittelten Werten für 400-600 nm (erste drei Zeilen in Tabelle 5.2) zeigt, dass mit der Einschränkung des Wellenlängenbereiches durchgehend etwas höhere Schichtdicken von 0,04 bis 0,07 nm ermittelt werden. Der Einfluss der Abweichungen unterhalb von 400 nm ist also gering, wenn nur eine (effektive) Schichtdicke zu ermitteln ist. Auch zeigen die beiden anderen Modelle bei Ausdehnung des gesamten Fitbereiches eine zunehmende Annäherung an das einfache Modell, d.h. größere Werte für den Fehler (MSE).

Als Beispiel hierfür ist für die Messung nach der HF-Behandlung der Unterschied der Ergebnisse in Abhängigkeit von der Wahl des Fitbereiches für das Modell mit der amorphen Zwischenschicht in Abbildung 5.5 dargestellt. So kann man feststellen, dass der Einfluss der wahrscheinlichen Messungenauigkeiten unterhalb von etwa 380 nm für Schichtdickenbestimmungen zwar sehr gering, aber wohl entscheidend ist für eine genauere Bewertung der Spektren hinsichtlich Oberflächenmodifikationen.

Insgesamt ist festzustellen, dass bei den dünnsten Schichten prinzipielle Schwierigkeiten bei der eindeutigen Auswertung der Spektren bestehen, und damit eindeutige Unterscheidungen bezüglich der Gültigkeit von verschiedenen Modellen der tatsächlich vorliegenden Struktur problematisch sind. Wie oben gezeigt, ist dies insbesondere dann der Fall, wenn nur ein beschränkter Messbereich oberhalb des Bereichs von 250 – 380 nm in dem sich Veränderungen an der Oberfläche von Silicium besonders empfindlich bemerkbar machen, zur Verfügung steht.

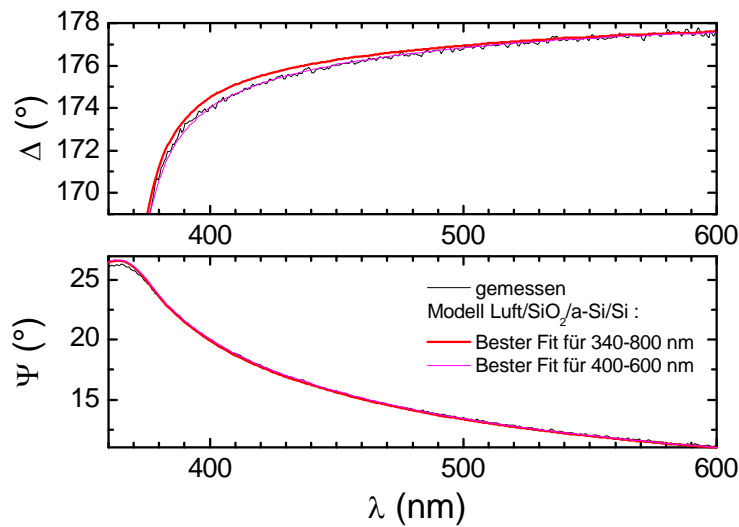


Abbildung 5.5: Vergleich der besten Fits bei Verwendung des Modells mit einer amorphen Zwischenschicht, bestimmt unter Berücksichtigung von unterschiedlichen Wellenlängenbereichen

5.1.5 Schlussfolgerungen

Aus den ellipsometrischen Messungen an Wafern, die entsprechend den zu dotierenden vorbehandelt wurden, ergeben sich folgende Erkenntnisse:

- Für eine Interpretation der ellipsometrischen Spektren kann die Struktur der Oberfläche mit einem einfachen Modell einer dünnen Oxidschicht gut angenähert werden.
- Die Oberflächen nach den verschiedenen chemischen Behandlungsschritten entsprechen zwar nicht der idealen sauerstofffreien wasserstoffterminierten Oberfläche, sie sind jedoch auch nicht stark gegenüber diesen idealen Annahmen verändert.
- Die Oberflächenbehandlungen konnten reproduzierbar durchgeführt werden.
- Geringe Variationen in der Vorbehandlung, insbesondere auch nach dem Oxid-entfernungsschritt unmittelbar vor der Dotierung, sind nicht kritisch für die Oberflächenreproduzierbarkeit.
- Das Vorliegen einer Interface-Schicht an der Oberfläche mit einer Ausdehnung, die geringer ist als eine Monolage SiO_2 ist wahrscheinlich.
- Eine oxidfreie Oberfläche ist mit den chemischen Standard-Vorbehandlungen nicht herzustellen, und die Oxidschicht wächst nach Kontakt mit Reinraumluft weiter an.
- In der Reinraumumgebung ist dieses Wachstum für die kurzen Zeiten des Transports der gereinigten Wafer zur die Vakuumschleuse der Prozesskammer jedoch unerheblich.

Neben der Bestimmung der Auswirkungen der Vorbehandlungen stellen die Eingangsmessungen Referenzwerte für eine folgende Einschätzung der Oberflächenveränderungen durch den Dotierprozess dar (siehe Kapitel 5.4.7).

5.2 Temperaturkalibrierungen

5.2.1 Zusammenhang von Temperatur und Lampenleistung

Die Kenntnis der genauen Wafertemperatur während der Kurzzeitprozesse ist für eine Analyse und Bewertung der Dotierergebnisse von entscheidender Bedeutung. Die Bestimmung und Kontrolle der Temperatur stellt bei Kurzzeitprozessen eine besondere Herausforderung dar.

Ein von der Temperaturerfassung unabhängiger und zusätzlicher Parameter als Hinweis auf die im Reaktor herrschenden Temperaturen ist die für den jeweiligen Prozess aufzuwendende Leistung. Zur Veranschaulichung des Zusammenhangs von Ist- und Sollwert der Temperatur und der für den gewählten Temperaturverlauf aufzubringenden Gesamtleistung der Lampen ist für einen neuen PinTC in der Abbildung 5.6 dargestellt.

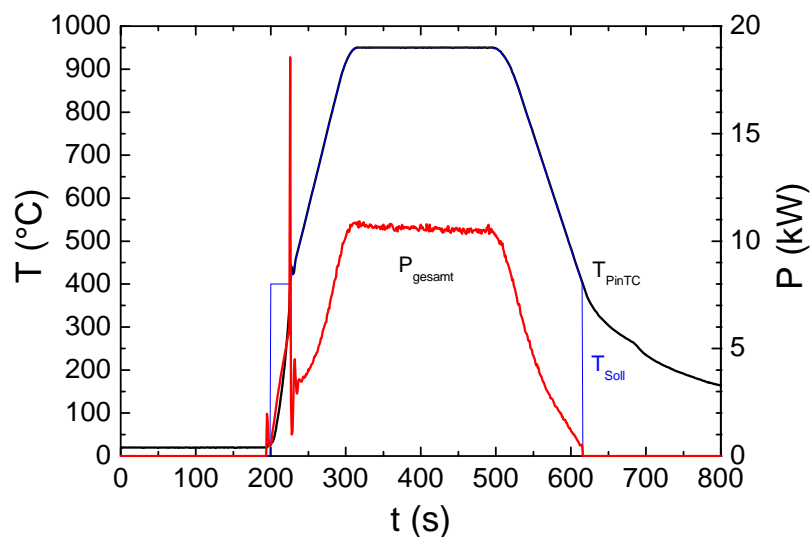


Abbildung 5.6: Zusammenhang von vorgegebenem Temperatursollwert, dem mit PinTC gemessenen Istwert und der Leistungsaufnahme aller Lampen während eines typischen Prozesses unter Einsatz eines neuen PinTCs.

In dieser Darstellung ist sehr gut zu erkennen, dass die Regelung der Lampenleistung gewährleistet, dass der mittels PinTC gemessene Temperaturwert fast exakt dem vorgegebenen Sollwert entspricht. Die Lampenleistung zeigt nur in einem Bereich des Umschaltens von unregelter kurzer Aufheizphase mit geringer Lampenheizung, dargestellt durch die konstante Temperaturvorgabe, ein kurzes Einschwingverhalten, das sich jedoch kaum in Temperaturschwankungen des PinTC bemerkbar macht. Im Bereich der Rampe und beim Umschalten auf die Prozesstemperatur treten keine besonderen Schwankungen der notwendigen Lampenleistung auf, und die Leistung ist stabil für den Bereich der konstanten Temperatur. Somit ist offensichtlich bei diesem Prozess eine ausreichend schnelle und auch stabile Temperaturmessung gegeben.

Alterung

Bei den anfänglichen Versuchen kam es jedoch zum wiederholten Bruch des Thermoelements des Original PinTC, ohne dass dies vorher in Auffälligkeiten des gemessenen Temperaturverlaufs deutlich wurde. Der Ausfall erfolgte jeweils nach Ende eines Prozesses während des Abkühlens. Sichtbare Spuren einer äußeren mechanischen Zerstörung konnten dabei nicht

gefunden werden. Ein Grund für diese Ausfälle war vermutlich eine schnelle Alterung des PinTC in den anfangs verwendeten wasserstoffreichen Prozessatmosphären bei sehr hohen Temperaturen.

In der Abbildung 5.7 sind drei Temperaturverläufe und die zugehörigen Lampenleistungen für vergleichbare Prozesse unter Verwendung desselben PinTC dargestellt, wobei der Prozess unmittelbar vor Ausfall des PinTC zwei vorangegangenen Prozessen gegenübergestellt wurde.

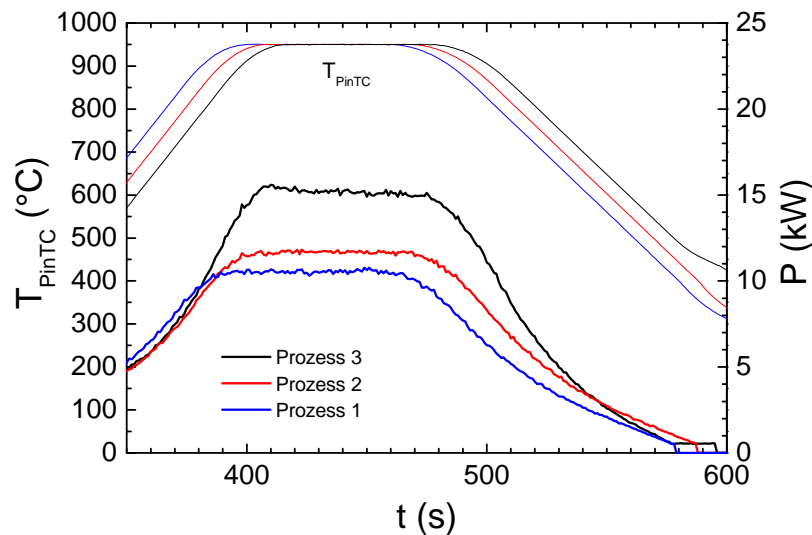


Abbildung 5.7: Temperaturverläufe und entsprechende Lampenleistungen vor dem Ausfall eines PinTC unmittelbar nach Prozess 3.

Es ist hier gut zu erkennen, dass anhand des gemessenen und geregelten Temperaturverlaufes keinerlei Abweichungen festzustellen waren, die Regelung arbeitete offenbar korrekt. Aufgrund der erheblich höheren Leistung für die (scheinbar) gleiche Temperatur vor dem Ausfall muss damit von einem Fehler der Temperaturmessung selbst ausgegangen werden. Dies bewirkte, dass mit einer jeweils höheren angeforderten Leistung tatsächlich eine höhere Temperatur einherging, die schließlich zu einer sich noch verstärkenden Alterung und zum letztendlichen Ausfall des Thermoelements führte.

Geringe Leistungsunterschiede von Prozess zu Prozess aufgrund unterschiedlicher Randbedingungen wie Kammerwandtemperatur, Gaszusammensetzung, zunehmender Kontamination von Kammerwand und Quarzglas oder auch leicht veränderter Lage des Wafers auf dem PinTC durch geringe Abweichungen bei der Beladung sind zu erwarten. Um diese zu berücksichtigen, ist eine Regelung mit geschlossenem Regelkreis (closed-loop) erforderlich. Die in der Abbildung dargestellten deutlich erhöhten Leistungen sind hingegen offenbar auf die Alterung des PinTC zurückzuführen.

Einfluss der Ausführung des Pin-Thermoelements

Nach den Ausfällen des PinTC wurden, wie eingangs beschrieben, Ersatzkonstruktionen benutzt. Es wurde eine freie Messstelle in direktem Kontakt zur Waferunterseite eingesetzt sowie eine mit Keramikzement verkapselte Messstelle. Der Vorteil bei letzterer war, dass kein direkter Metallkontakt zum Wafer bestand und durch den keramikartigen Überzug die Messstelle fixiert und mechanisch stabiler war. Nachteilig war, dass dieser Aufbau eine größere thermische Trägheit besaß und daher Abweichungen bei der Bestimmung der exakten Temperatur des Wafers zu erwarten waren.

Anhand von Leistungskurven von typischen Prozessverläufen soll dies illustriert werden. In Abbildung 5.8 sind für verschiedene Prozesse mit jeweils 900 °C Soll-Haltetemperatur die zugehörigen Werte der aufgewendeten Lampenleistungen dargestellt.

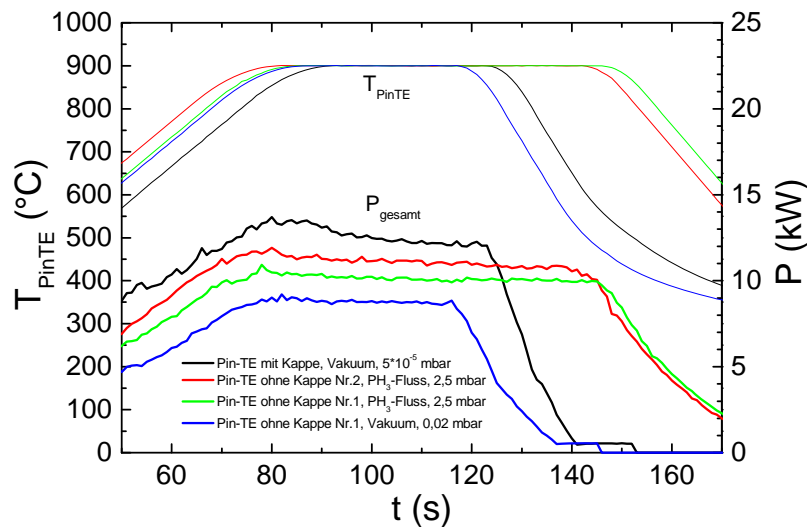


Abbildung 5.8: Vergleich der Gesamtleistungsaufnahme bei verschiedenen Prozessen mit 900°C Soll-Haltetemperatur unter Einsatz verschiedener Pin-Thermoelemente.

Man kann hier erkennen, dass sich die Form der Leistungskurven bei den unverkapselten Thermoelementen kaum von der des Original-PinTCs unterscheidet; die hier eingesetzten Messstellen zeigen somit ein ähnliches Ansprechverhalten. Wie auch bereits bei dem Original-PinTC beobachtet wurde, zeigt der Vergleich der Kurven für dasselbe Pin-Thermoelement bei unterschiedlichen Prozessbedingungen (ohne Gasfluss-blau und mit Dotiergasfluss-grün) einen Unterschied der Leistungsaufnahme, was belegt, dass die verschiedenen Prozessbedingungen auch andere Heizleistungen erforderlich machen. So ist bei stationärem Gasfluss offenbar ein erhöhter Heizbedarf erforderlich. Auch ist bei gleichem Prozess und verschiedenen Pin-Thermoelementen gleicher Ausführung (nach Wechsel des Thermoelements ohne Verkapselung, Kurven rot und grün) eine Verschiebung der Leistungskurve deutlich. Als ein Hauptgrund hierfür wurde eine durch den Ein- und Ausbau des Pin-Thermoelements bedingte Verschiebung der Lage der Messstelle bezüglich des Waferrandes festgestellt. Die durch die Konstruktion bedingte Anordnung des Pin-Thermoelements nahe dem Waferrand führte zu Messfehlern bei nur geringfügiger Verschiebung der Messposition.

Eine deutliche Verschiebung der Leistungskurve ist jedoch nach Einbau des Thermoelements mit der Oxidverkapselung festzustellen. Der höhere Leistungsbedarf besonders auch nach Beendigung der Aufheizrampe kann auf eine höhere thermische Trägheit des Systems zurückgeführt werden, wie auch auf die gegenüber dem unverkapselten Pin-Thermoelement veränderte thermische Umgebung der Messstelle mit unterschiedlichem Absorptions- und Emissionsverhalten. Eine dadurch gemessene scheinbar geringere Wafertemperatur führte zu einem erhöhten geforderten Heizbedarf durch die Lampen.

Einfluss der Aufheizrampe

Ein Einfluss der Steilheit der Aufheizrampe auf die erforderliche Lampenleistung war bei der Verwendung des Thermoelementes ohne Verkapselung nachzuweisen, wie in der Abbildung 5.9 gezeigt ist.

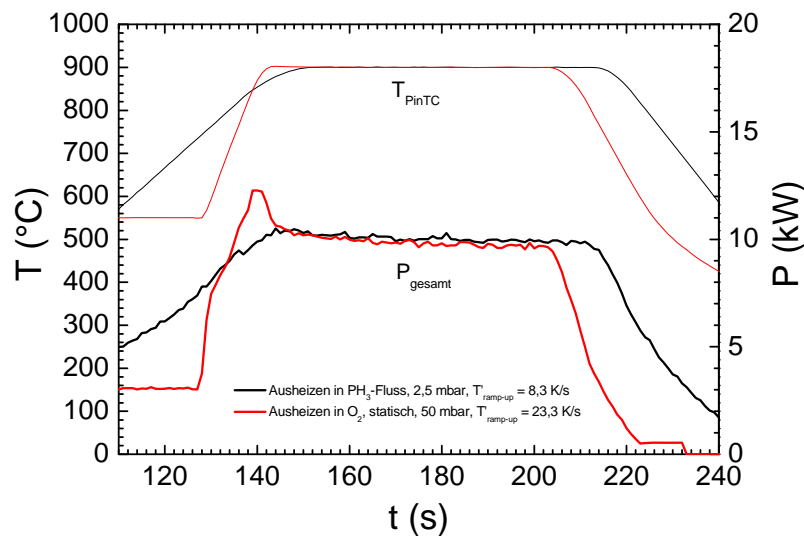


Abbildung 5.9: Vergleich der Leistungsaufnahme bei verschiedenen Heizrampen auf 900 °C Solltemperatur bei Verwendung eines Pin-Thermoelements ohne Verkapselung.

Hier ist ein deutlich erhöhter Leistungsbedarf beim Aufheizen auf 900 °C Solltemperatur bei Durchführung einer steileren Rampe von ca. 25 K/s (rot) gegenüber der der von ca. 10 K/s (schwarz) zu erkennen.

Reproduzierbarkeit

Der Einsatz des verkapselten Thermoelements, das für den Großteil der Dotierversuche eingesetzt wurde, erwies sich günstig für die Reproduzierbarkeit der Prozesse. In der Abbildung 5.10 sind zwei Temperatur- und zugehörige Lampenleistungsverläufe von Prozessen dargestellt, die unter denselben Prozessbedingungen durchgeführt wurden, zwischen denen jedoch 24 verschiedene Dotierprozesse durchgeführt wurden. Ein Unterschied zwischen dem erforderlichen Heizleistungsbedarf ist hier nicht zu verzeichnen.

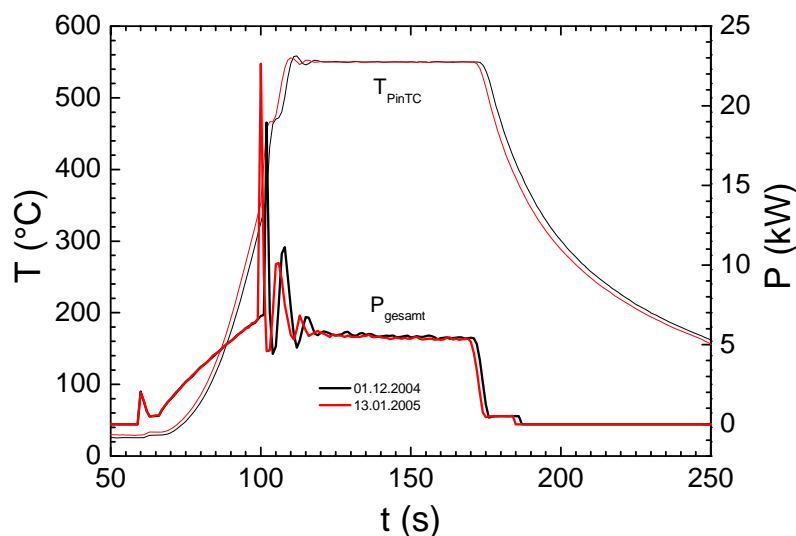


Abbildung 5.10: Vergleich der Reproduzierbarkeit der benötigten Lampenleistung bei Verwendung des Pin-Thermoelements mit Verkapselung.

5.2.2 Temperaturkalibrierungen über mit Thermoelement bestücktem Wafer

Bei der Messung der Wafertemperatur mittels eines TC-Wafers wurde deutlich, dass diese erheblich über der am Pin-Thermoelement gemessenen lag. Dies war besonders für das Pin-Thermoelement mit der Oxidkappe bei schnellen Aufheizrampen der Fall, wie er bei den meisten der Dotierexperimente vorlag und im Folgenden dargestellt ist. In der Abbildung 5.11 sind Aufzeichnungen der Temperatur des TC-Wafers und der zugehörigen Lampenleistungen der Testprozesse dargestellt, bei denen über die Temperaturmessung am Pin-Thermoelement und einer entsprechenden Regelung der Lampenleistung die vorgegebenen Solltemperaturen eingestellt wurden. Die Verläufe der Temperaturen des Pin-Thermoelements sind hier nicht mit dargestellt, sie entsprechen fast exakt den Solltemperaturen (siehe Abbildung 5.6). Zusätzlich sind in der Abbildung 5.11 die aufgewendeten Lampenleistungen bei drei Dotierprozessen bei jeweils verschiedenen Solltemperaturen dargestellt.

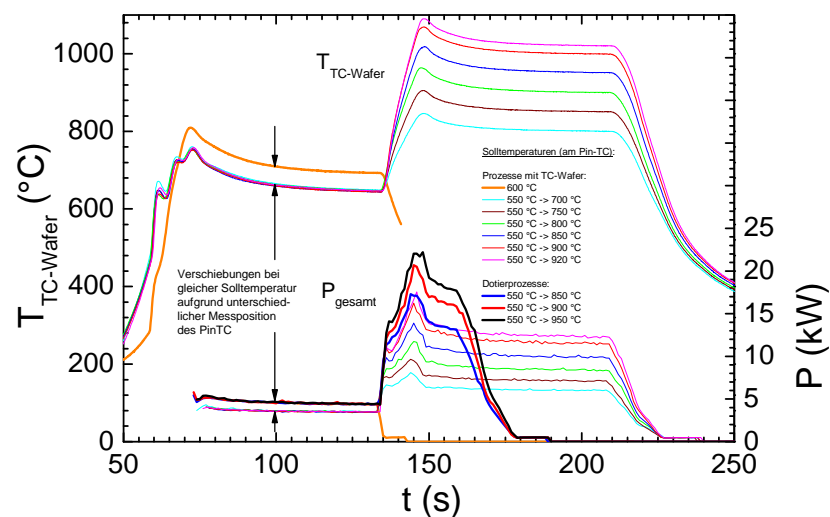


Abbildung 5.11: Darstellung der Temperaturverläufe, gemessen in der Wafermitte mittels TC-Wafer (obere Kurven), und der zugehörigen Verläufe der Gesamtleistung der Lampen (untere Kurven) für verschiedene Testprozesse im Vergleich zu Leistungsverläufen dreier Dotierprozesse. Die jeweiligen Solltemperaturen, die den mittels Pin-Thermoelement gemessenen Temperaturen entsprechen, sind angegeben.

Aus dieser Abbildung ist zu ersehen, dass eine Überhöhung der Lampenleistung nach einer schnellen Aufheizrampe mit einer Überhöhung der Wafertemperatur einhergeht, wie bereits zur Abbildung 5.9 diskutiert wurde. Wie dort zu ersehen war, ist diese Überhöhung in dem Verlauf der Temperatur des Pin-Thermoelements nicht ausgeprägt. Vielmehr ist dessen thermische Trägheit der eigentliche Grund dafür, dass die Lampenleistung entsprechend der dadurch geringeren Temperatur für die Einhaltung des Sollwertes überhöht worden ist.

Außerdem zeigt sich im Vergleich der Prozessverläufe bei den Testprozessen mit denen bei der Dotierung, dass bei letzteren eine deutlich höhere Lampenleistung für das Erreichen einer gleichen vorgegebenen Solltemperatur bei gleicher Form der Leistungskurve erforderlich war. Diese Verschiebung der Leistungskurve ist in der Abbildung 5.11 gekennzeichnet. Die benötigte Leistung für den Vorheizschritt der Dotierung bei 550 °C Solltemperatur ist höher als die entsprechende Leistung beim TC-Wafer-Prozess. Die gleiche Leistung wird beim Testprozess erst beim Einstellen einer Solltemperatur von 600 °C aufgewendet (orange Linie). Die Verschiebung bei den Dotiertemperaturen selbst ist durch direkten Vergleich entsprechender Leistungskurven (Solltemperatur: 850 °C bei Dotierung - blau, 920 °C mit TC-Wafer - magenta) ersichtlich.

Diese deutliche Verschiebung kann nur dadurch erklärt werden, dass das Pin-Thermoelement bei den Dotierprozessen aufgrund der dafür erfolgten automatischen Beladung näher am Waferrand lag als beim manuell eingelegten TC-Wafer (siehe Abbildung 4.5). Eine Messposition sehr nahe am Waferrand führt offenbar zu deutlichen Messfehlern bei der Temperaturbestimmung. Ursache dafür können erhöhte Abstrahlungsverluste am Rand des Prozesswafers sein, die dort im Gleichgewichtsfall (steady state) von Absorption und Emission zu niedrigeren Temperaturen führen¹. Eine Möglichkeit, diesen Effekt zu mindern, ist der Einsatz eines Siliciumringes (Guardring) [Nenyei '93], der den Prozesswafer umschließt und der damit gewissermaßen den äußeren Rand der beheizten Siliciumfläche bildet. Die Waferaufnahme der Anlage ließ sich jedoch nicht mit einem Guardring ausstatten.

Für eine Abschätzung der wahren Wafertemperatur bei den Dotierprozessen wurde nun eine Kalibrierung über die aufgezeichneten Lampenleistung durchgeführt. Der Leistungsbedarf für das Erreichen verschiedener Solltemperaturen bei Testprozessen, für die die genaue Wafertemperatur mittels der TC-Wafer-Messungen bekannt war, wurde ermittelt. In der Abbildung 5.12 sind die bei diesen Testprozessen ermittelten Temperaturverläufe auf der Wafermitte während der Aufheizrampen und für die Haltezeit bei der Solltemperatur dargestellt und die gewählten Vergleichstemperaturen am Beispiel des 920 °C-Testprozesses eingetragen. Dies sind die Maximaltemperatur T_{max} als die für das Dotierergebnis kritische Temperatur, die Temperatur T_{10s} zum Vergleich mit der 10 s nach der Aufheizrampe erforderlichen Lampenleistung² und außerdem noch die Temperatur T_{60s} für den stabilisierten Zustand nach 60 s Haltezeit bei der Solltemperatur.

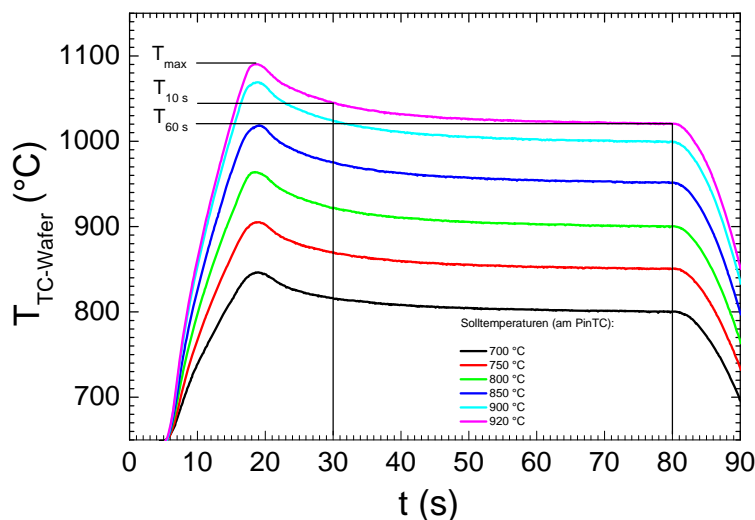


Abbildung 5.12: Darstellung der Temperaturverläufe, gemessen in der Wafermitte mittels TC-Wafer, bei jeweils angegebener konstanter Solltemperatur (für 60 s Haltezeit). Für den Testprozess bei 920 °C Solltemperatur ist die Bestimmung der Maximaltemperatur (T_{max}), der Temperatur T_{10s} (entsprechend dem Ende eines 10 s Dotierprozesses) und der Temperatur T_{60s} angegeben.

¹ Aus Untersuchungen des konventionellen optischen Heizens ist bekannt, dass sich die Waferoberfläche mit verschiedenen Heizraten aufheizen kann. Während in der Aufheizphase durch den so genannten „photon box effect“ eine lokale Überhöhung der Temperatur am Waferrand auftreten kann, wurde bei Gleichgewicht von Absorption und Emission eine erhöhte Wärmeabstrahlung und konvektive Kühlung am Waferrand beobachtet (siehe [Nenyei '93]).

² Ein Vergleich der Lampenleistungen 10 s nach der Aufheizrampe ist günstiger als der Vergleich der maximalen Leistungen, da hier geringere Schwankungen auftreten, wie die Verläufe der Lampenleistungen in Abbildung 5.11 zeigen.

Diese Werte für die verschiedenen Testprozesse sind zusammen mit den Leistungen in der Abbildung 5.13 dargestellt. Die jeweiligen Temperaturwerte können gut durch eine lineare Regression beschrieben werden, für die jeweiligen Leistungen erwies sich ein Polynom zweiter Ordnung als günstiger Fit. Damit zeigt die Wafertemperatur einen eindeutigen Zusammenhang zur Lampenleistung. Über diesen Zusammenhang lassen sich nun auch die Temperaturen bei den Dotierprozessen angeben. In der Abbildung 5.13 sind die Leistungen bei den Dotierprozessen beim Ende der 10 s Solltemperatur-Haltezeit eingetragen. Über die bei den entsprechenden Testprozessen notwendige Leistung kann somit auf die jeweiligen Temperaturen geschlossen werden.

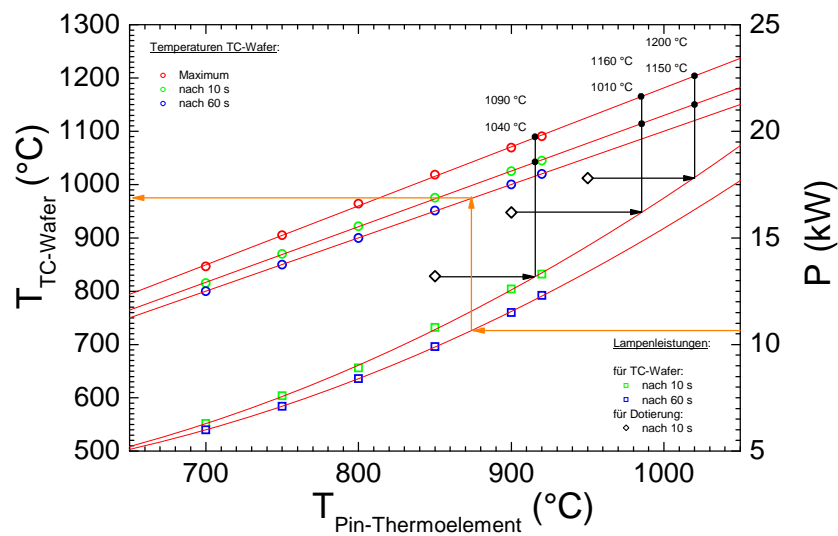


Abbildung 5.13: Darstellung eines Beispiels der Temperaturermittlung über die aufgewendete Lampenleistung. Angegeben sind die mittels TC-Wafer gemessenen Temperaturen und die bei diesen Testprozessen benötigten Lampenleistungen für verschiedene Prozesszeiten (nach 10 s und nach 60 s) in Abhängigkeit von der eingestellten Solltemperatur (Temperatur des Pin-Thermoelements). Zum Vergleich sind die bei drei Dotierprozessen benötigten Lampenleistungen angegeben (nach 10 s Haltezeit) und die daraus zu erwartenden tatsächlichen Temperaturen (schwarze Pfeile) sowie die Leistung eines Oxidationsprozesses bei einer Solltemperatur von 950 °C unter Verwendung eines Original-PinTCs (orange Pfeile).

Zum Vergleich wurde der Leistungsbedarf bei einem Oxidationsprozess bei 950 °C unter Verwendung eines neuen Original-PinTCs (vor dessen Ausfall, siehe Abbildung 5.6) in diese Abbildung eingetragen. Eine entsprechende Ermittlung der Wafertemperatur liefert hier 975 °C¹. Der Original-PinTC zeigte offenbar nicht die hohe thermische Trägheit des Ersatzes, wie oben bereits beschrieben. Andererseits kann durch diesen Vergleich festgestellt werden, dass eine Temperaturkalibrierung über die Leistungsdaten erfolgen kann. Eine Zusammenfassung ausgewählter, aus den Leistungsdaten ermittelter Dotiertemperaturen ist in der Tabelle 5.3 angegeben.

Tabelle 5.3: Ermittelte Wafertemperaturen und Aufheizrampen von Dotierprozessen (Vorheizen 60 s, Eintreiben 10 s).

Solltemperatur	Sollrampe	Istrampe	T_{max}	T_{10s}	T_{60s}
550 °C			810 °C	740 °C	690 °C
850 °C	20 K/s	32 K/s	1090 °C	1040 °C	-
900 °C	23 K/s	38 K/s	1160 °C	1110 °C	-
950 °C	27 K/s	41 K/s	1200 °C	1150 °C	-

¹ Die im Vergleich geringe Differenz zur Solltemperatur ergibt sich hier wahrscheinlich aufgrund des Randeffektes, da die Messposition des Original-PinTC auch nahe des Waferrandes lag (siehe Abbildung 4.5).

Anhand der in dieser Tabelle angegebenen Vergleiche von Soll- und Isttemperaturen ist abzulesen, dass die bei den Kurzzeitprozessen erzielten Wafertemperaturen deutlich höher sind als die laut Rezept geforderten. Bei der Auswertung der Dotierprofile ist dies zu berücksichtigen.

5.2.3 Untersuchungen zur Homogenität der Temperaturverteilung der Anlage

Eine Abschätzung der Temperaturverteilung auf den Wafern während der Kurzzeitprozesse erfolgte durch Kurzzeitoxidationen von gereinigten Wafern bei vorgegebenen Temperaturen in Sauerstoffatmosphären. Die Dicke der Oxidschicht ist dann im Wesentlichen abhängig von der Temperatur auf dem Wafer. Je höher die Temperatur ist, umso größer ist die Schichtdicke; Temperaturinhomogenitäten auf dem Wafer während des Prozesses bilden sich somit über Schichtdickeninhomogenitäten der Oxidschicht ab. Die Ergebnisse von Oxidationen bei einem Sauerstoffdruck von 40 mbar ohne Gaszufuhr während der Heizphasen und von Oxidationen unter strömendem Sauerstoff bei 5 mbar stationärem Druck wurden ausgewertet und verglichen.

Statische Oxidation

Eine Darstellung der Schichtdickenverteilung nach einer statischen Oxidation (40 mbar O₂, 60 s bei 550 °C und 60 s bei 900 °C, jeweils Solltemperaturen) ohne Korrektur der Lampenleistungen zeigt die Abbildung 5.14. Die Schichtdicken sind hier zur qualitativen Bewertung in einem plastischen Höhenmodell abgebildet, die Lage des Waferflats ist in der Abbildung angegeben. Bei allen Homogenitätsuntersuchungen wurde der Waferflat so wie bei den meisten Dotierexperimenten parallel zu den Lampenstäben in Richtung des Beladeports und damit der Gaseinströmung ausgerichtet.

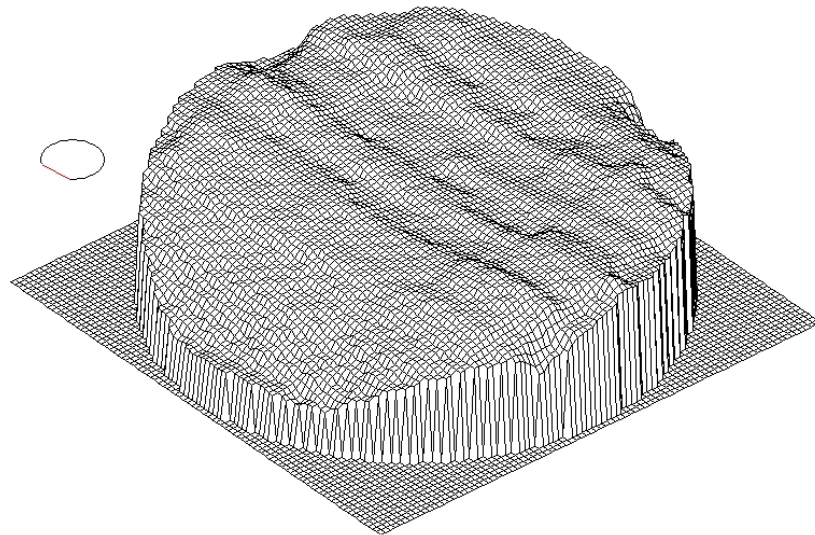


Abbildung 5.14: Oxidschichtdickendarstellung im Höhengitternetz eines Wafers nach statischer Oxidation ohne Leistungskorrektur bei 900°C Solltemperatur.

In der Abbildung sind wellenförmige Schichtdickenschwankungen zu erkennen, deren Breite sich über den gesamten Wafer erstreckt. Außerdem ist ein Abfallen der Schichtdicke zum Flat hin auffällig. Eine günstigere quantitative Auswertung dieser Beobachtungen bietet eine Schnittdarstellung entlang des Waferdurchmessers im rechten Winkel zum Flat. Nur entlang dieser Richtung ist wegen der parallelen Anordnung der Stablampen eine Beeinflussung der Homogenität durch Wichtung der Einzellampenleistungen möglich. Ein solcher „line scan“ ist

in der folgenden Abbildung dargestellt. Zu beachten ist, dass sich die dort und in den folgenden Abbildungen mit angegebenen statistischen Ergebnisse der Schichtdickenauswertungen jeweils auf den gesamten Wafer beziehen.

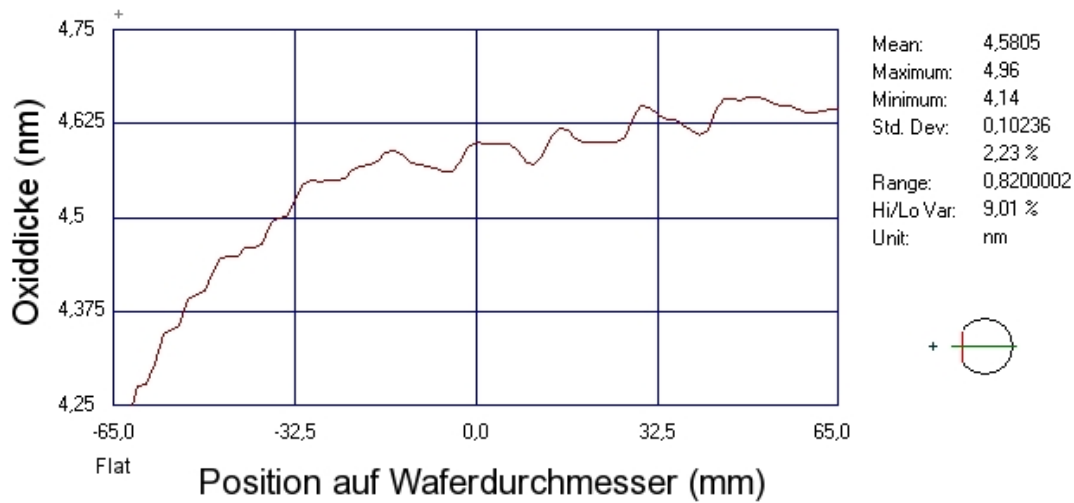


Abbildung 5.15: Schichtdickenverlauf bei statischer Oxidation des Wafers aus Abbildung 5.14 entlang der rechts auf dem Wafersymbol dargestellten Messlinie.

Die Abnahme der Schichtdicke zum Flat hin und die wellenförmige Struktur zeichnen sich hier deutlich ab. Im Messbereich von 130 mm Durchmesser ist ein Abfall der Schichtdicke von ca. 0,4 nm zum Flat hin deutlich.

Stationäre Oxidation

Bei der stationären Oxidation, die auch den Druck und Strömungsverhältnissen der meisten Dotierversuche entspricht, wurden deutlich geringere Schichtdicken aufgrund des geringeren Sauerstoffdruckes erzielt als bei der statischen Oxidation. In der Abbildung 5.16 ist ein entsprechendes Schichtdickenprofil dargestellt. Dabei spielte die Gasströmung selbst nur eine untergeordnete Rolle, eine statische Oxidation bei nur 5 mbar ergab Schichtdicken im gleichen Bereich.

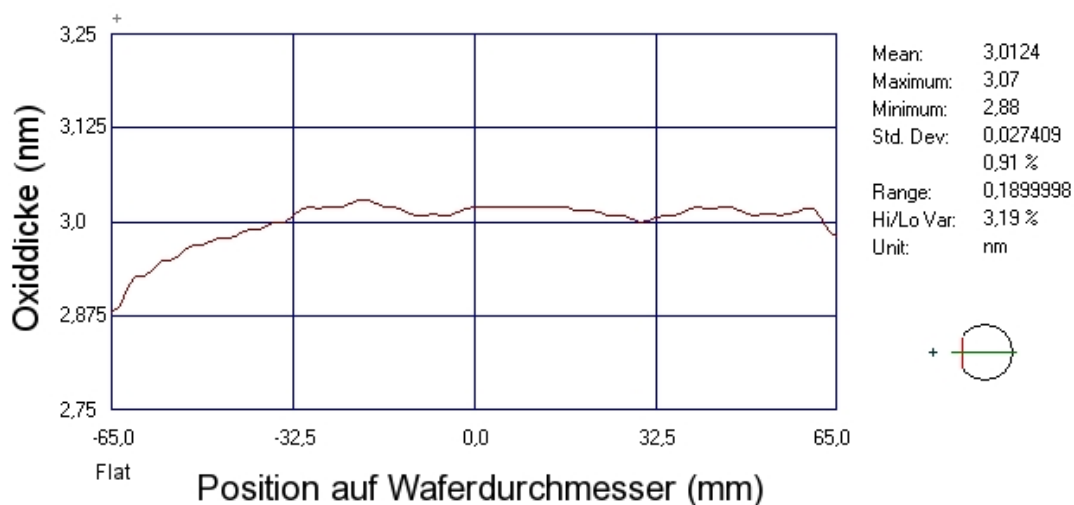


Abbildung 5.16: Schichtdickenverlauf bei stationärer Oxidation bei 5 mbar mit gleicher Leistung wie beim Wafer aus Abbildung 5.15 entlang der rechts auf dem Wafersymbol dargestellten Messlinie.

Die Abhängigkeit der Schichtdicke von den Prozesstemperaturen ist hier geringer, die Schichtdickenschwankungen sind damit deutlich geringer als bei der Oxidation mit höherem Druck. Somit ist diese Prozessführung kaum für ein Monitoring der Temperaturverteilung geeignet.

Leistungskorrektur

Um nun auch mit der statischen Oxidation bei einem höheren Druck von 40 mbar eine möglichst gleichmäßig dicke Oxidschicht zu erhalten, wurde versucht, die Leistungswerte der Lampen entsprechend anzupassen. Die Ergebnisse verschiedener Korrekturen wurden über die Auswertung der Schichtdickentopographien verglichen.

Eine nur grobe Anpassung über die Wichtung der Gesamtleistungen der drei Lampengruppen durch eine Veränderung von Rezeptparametern erwies sich hierfür als wenig erfolgreich. Es wurden daraufhin die Leistungen der Einzellampen gewichtet, was über das Einstellen separater Wichtungsfaktoren in einer Lampenkorrekturtabelle möglich ist. Bei der Bestimmung der Soll-Leistungen der Lampen in einer Berechnung durch das Regelsystem werden diese berücksichtigt. Ein Beispiel für solche Korrekturwerte der jeweiligen Lampen ist in der Abbildung 5.17 angegeben.

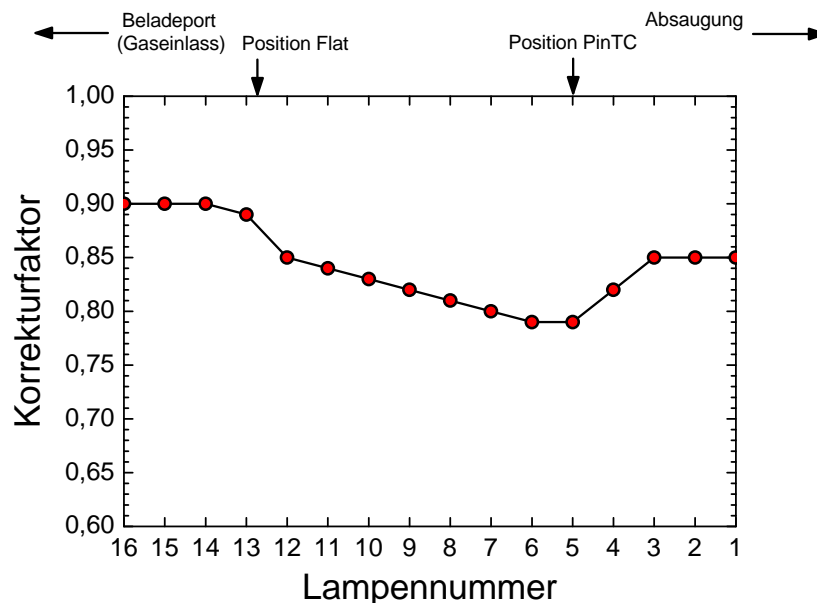


Abbildung 5.17: Darstellung der Korrekturfaktoren der Lampen der bezeichneten Position für eine verbesserte Schichtdickenverteilung.

Dargestellt sind hier die Lampennummern mit der jeweiligen Position bezüglich der Kammerbreite und die entsprechenden Wichtungsfaktoren. Ein Wert von 1 bedeutet volle Leistung, 0 keine Leistung.

Die damit erzielten Schichtdickenprofile für statische und stationäre Oxidationen sind in den folgenden Abbildungen dargestellt.

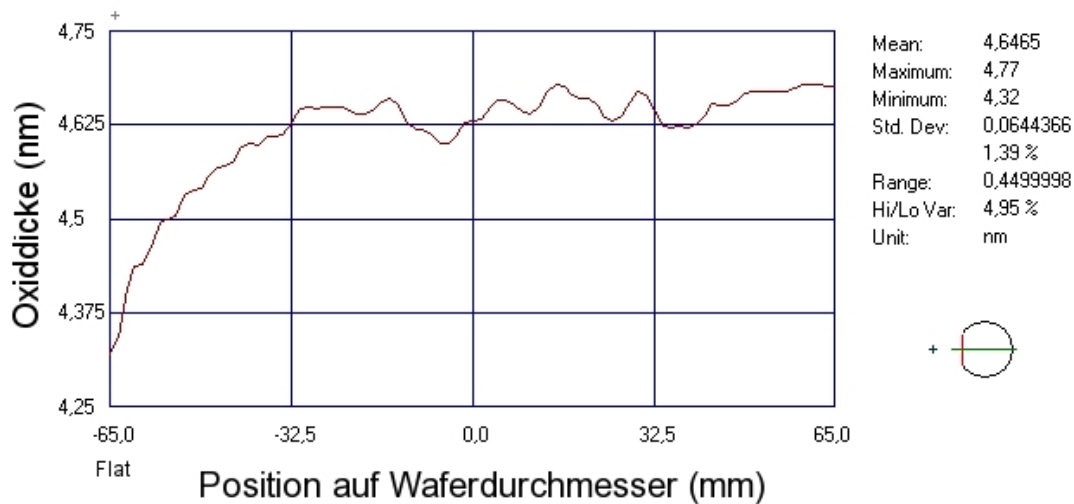


Abbildung 5.18: Schichtdickenverlauf bei statischer Oxidation mit Leistungskorrektur entlang der rechts auf dem Wafersymbol dargestellten Messlinie.

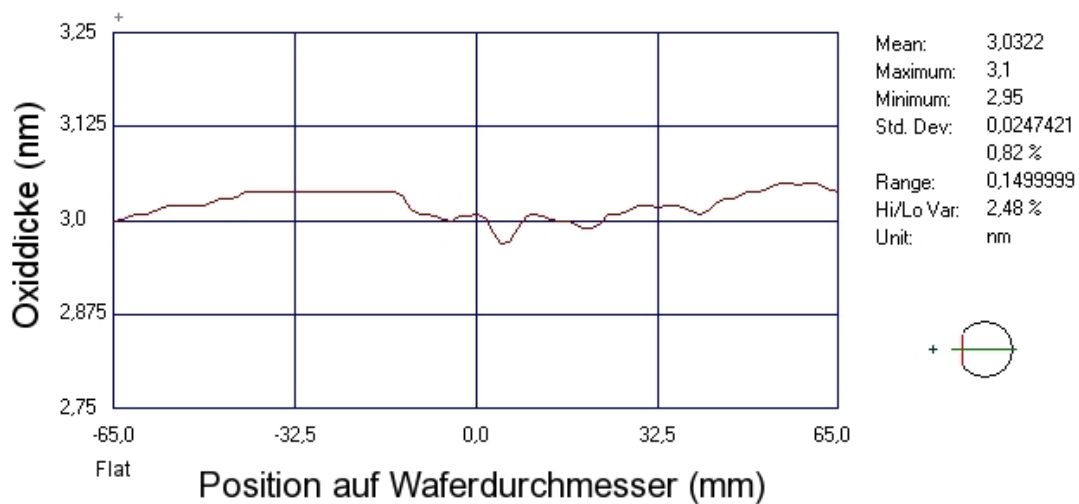


Abbildung 5.19: Schichtdickenverlauf bei stationärer Oxidation mit Leistungskorrektur entlang der rechts auf dem Wafersymbol dargestellten Messlinie.

Die Abbildung 5.18 zeigt im Vergleich zur Oxidation ohne Leistungskorrektur in Abbildung 5.15 eine etwas ebenere Schichtdickenverteilung auf der Wafersseite gegenüber dem Flat und eine zwar etwas geringer ausfallende, aber noch deutliche Verringerung der Schichtdicke zum Flat hin. Die wellenförmigen Schwankungen wurden nicht vermindert. Die Oxidation bei geringem Druck unter stationärer Strömung zeigt eine homogenere Schichtdickenverteilung auf dem gesamten Wafer, die verbleibenden Temperaturunterschiede wirken sich hier nicht so kritisch auf die Schichtdickenverteilung aus.

Im Rahmen der hier beschriebenen Untersuchungen konnte keine optimale Korrektur gefunden werden. Dafür sind umfangreichere Versuchsreihen mit einer systematischen Auswertung erforderlich [Tillmann '95]. Die konstruktiven Beschränkungen der Anlage begrenzen jedoch die Möglichkeiten einer weiteren deutlichen Verbesserung der Temperaturverteilung auf dem Wafer.

Zusammenhang von Schichtdicke und Temperatur

Da die Oxidation ein temperaturaktivierter Prozess ist, kann davon ausgegangen werden, dass eine geringere Schichtdicke auch eine geringere Temperatur während des Oxidationsprozesses anzeigt. Experimente mit verschiedenen Temperaturen bei der statischen Oxidation bei einem Druck von 40 mbar zeigten dies deutlich. Daraus sollte eine Abschätzung vorgenommen werden, welchen Temperaturunterschieden die ermittelten Schichtdickenschwankungen entsprechen.

Die Wafertemperaturen wurden, wie oben beschrieben, aus den Lampenleistungen ermittelt. Für eine genauere Vergleichbarkeit wurden stabilisierte Temperatur und Leistung nach 60 s Haltezeit gewählt (siehe Abbildung 5.12). Die Maximaltemperaturen mögen für die absolute Schichtdicke zwar von etwas größerer Bedeutung sein, jedoch stehen diese in festem Zusammenhang zur hier gewählten Vergleichstemperatur. An dieser Stelle soll es nicht um die Auswertung der Oxidationskinetik gehen, sondern nur um eine Abschätzung der Temperaturverteilung auf dem Wafer aus den Schichtdickenschwankungen, wofür dieser Temperaturreferenzwert T_{60s} genügt. Da für die Temperaturkalibrierung ein TC-Wafer mit der Messposition in der Mitte benutzt wurde, ist aus den Schichtdickentopographien auch die entsprechende Schichtdicke aus dem Zentrum des Wafers herangezogen worden. Dafür wurde der Mittelwert der Schichtdicke des Zentrumsbereiches von 20 mm Durchmesser gewählt. Die Abhängigkeit der Schichtdicke von der Wafertemperatur ist in der Abbildung 5.20 dargestellt.

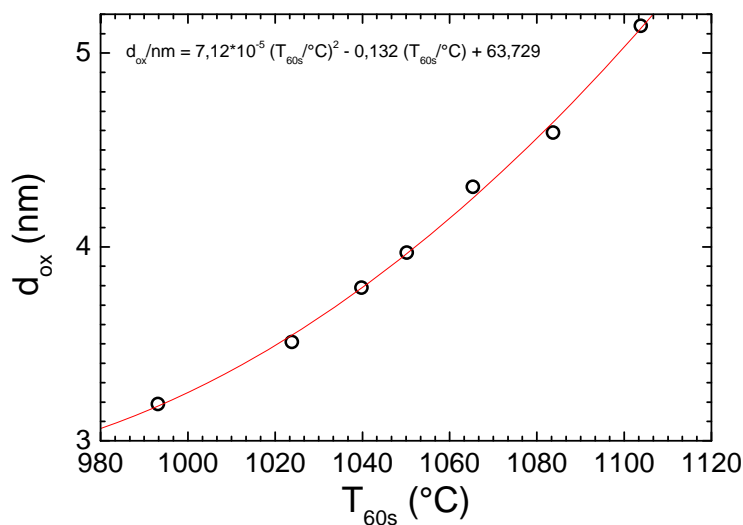


Abbildung 5.20: Oxidschichtdicke in Abhängigkeit von der Wafertemperatur für statische Oxidation bei 40 mbar ohne Leistungskorrektur.

Die rote Linie in dieser Abbildung stellt einen parabolischen Fit durch die Datenpunkte dar, dessen Formel mit angegeben ist.

Unter der Annahme, dass die Schichtdicke hauptsächlich von der Prozesstemperatur während der Haltezeit von 60 s abhängig ist, können mit Hilfe dieses ermittelten parabolischen Zusammenhangs zwischen Temperatur und Schichtdicke die Schichtdickenwerte in entsprechende Wafertemperaturwerte umgerechnet werden. Aus dem in Abbildung 5.20 dargestellten Schichtdickenverlauf ergibt sich damit der in Abbildung 5.21 angegebene Temperaturverlauf.

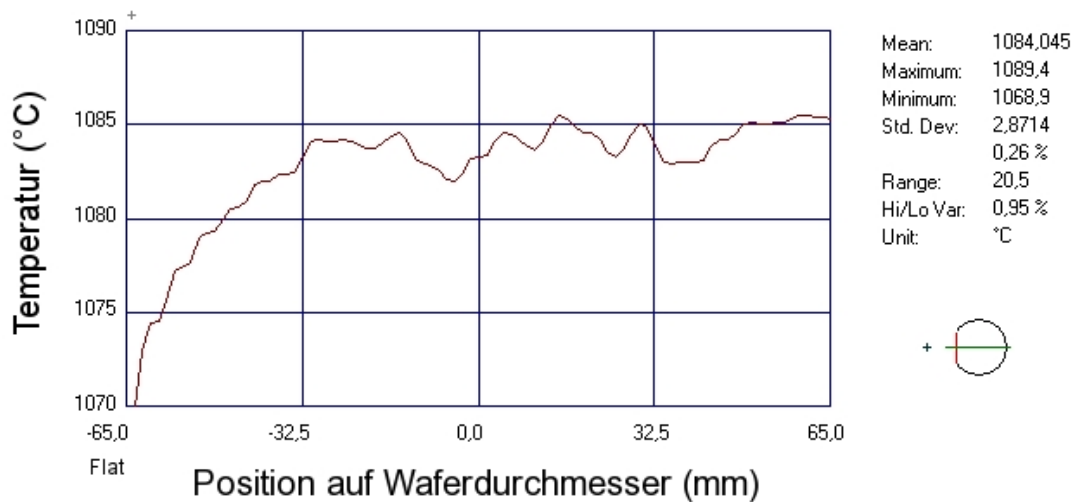


Abbildung 5.21: Temperaturverlauf bei statischer Oxidation mit Leistungskorrektur entlang der rechts dargestellten Messlinie, ermittelt aus dem Schichtdickenverlauf aus Abbildung 5.18 und dem in Abbildung 5.20 dargestellten Zusammenhang zwischen Schichtdicke und Wafertemperatur.

Insgesamt ergibt sich aus dieser Temperaturabschätzung bei diesem Wafer eine mittlere Temperatur von 1084 °C (bei einer Solltemperatur von 900 °C) mit einer Standardabweichung von 0,26 % für den untersuchten Bereich von 130 mm Durchmesser. Die maximale Temperaturschwankung betrug hier ca. 20 K. Angemerkt sei an dieser Stelle, dass aufgrund des Temperaturprofils während des Prozesses und der oben beschriebenen Methode der Temperaturermittlung die Maximalwerte der jeweiligen Temperaturen ca. 7,5 % oberhalb der hier ermittelten Werte lagen.

Entsprechend lässt sich nun auch eine Temperaturkarte des Wafers aus den ermittelten Schichtdickenwerten angeben. Für den Wafer aus Abbildung 5.21 (statische Oxidation bei 40 mbar Sauerstoffdruck, mit Leistungskorrektur) ergibt sich die in Abbildung 5.22 angegebene Darstellung.

Hier kennzeichnen die Farben den in der Legende angegebenen Temperaturbereich und die Linien stellen Isothermen im Abstand von 0,1% des Temperaturmittelwertes, eingezeichnet als grüne Linie, dar. Deutlich sind der durch die Leistungskorrektur erzielte Bereich gleichmäßigerer Temperaturverteilung zu erkennen (gelb) und ein Abfall der Temperaturen zum Flat hin. Die Position der Stablampen bezüglich dieses Wafers ist parallel zum Flat, das heißt, die heißeren Bereiche links und rechts des Wafers lassen sich nicht über eine Leistungsanpassung korrigieren. Auch ist ein scheinbar heißerer Bereich etwas oberhalb der Mitte auszumachen, der bei den meisten Wafers deutlich wurde. Die Ursache dieser Temperaturinhomogenität konnte nicht ermittelt werden.

Besonders auffällig sind etwas kühlere Stellen an drei Randpositionen, die bei allen Wafers festgestellt wurden. Diese Stellen entsprechen den Positionen der Auflagepunkte des Wafers, die in der Abbildung als rote Kreise eingezeichnet sind, wobei die oben rechts dargestellte die Position des Pin-Thermoelementes ist. Eine Ursache für diesen Effekt kann eine Wärmeleitung vom Wafer auf die Stifte sein und darüber hinaus eine Abschattung vom Lampenlicht und damit der Wärmeeinstrahlung auf den Wafer durch die Stifte und das Thermoelement, da nur ein unteres Lampenfeld vorhanden war. Deutlich wird hier ein weiteres Problem der Temperaturmessung mittels des Pin-Thermoelementes. Das Element selbst vermindert die lokale Wafertemperatur durch Kontakt und Abschattung, was bei Verwendung als Messeinrichtung zur Bestimmung der Regelgröße zu einer erhöhten Leistungsanforderung und damit zu einer höheren Wafertemperatur führt. Der Auflagepunkt sollte möglichst kleine

Bereiche des Wafers thermisch beeinflussen und daher möglichst nahe am Rand liegen, andererseits ist jedoch hier eine genaue Temperaturmessung aufgrund der Randeffekte kritisch.

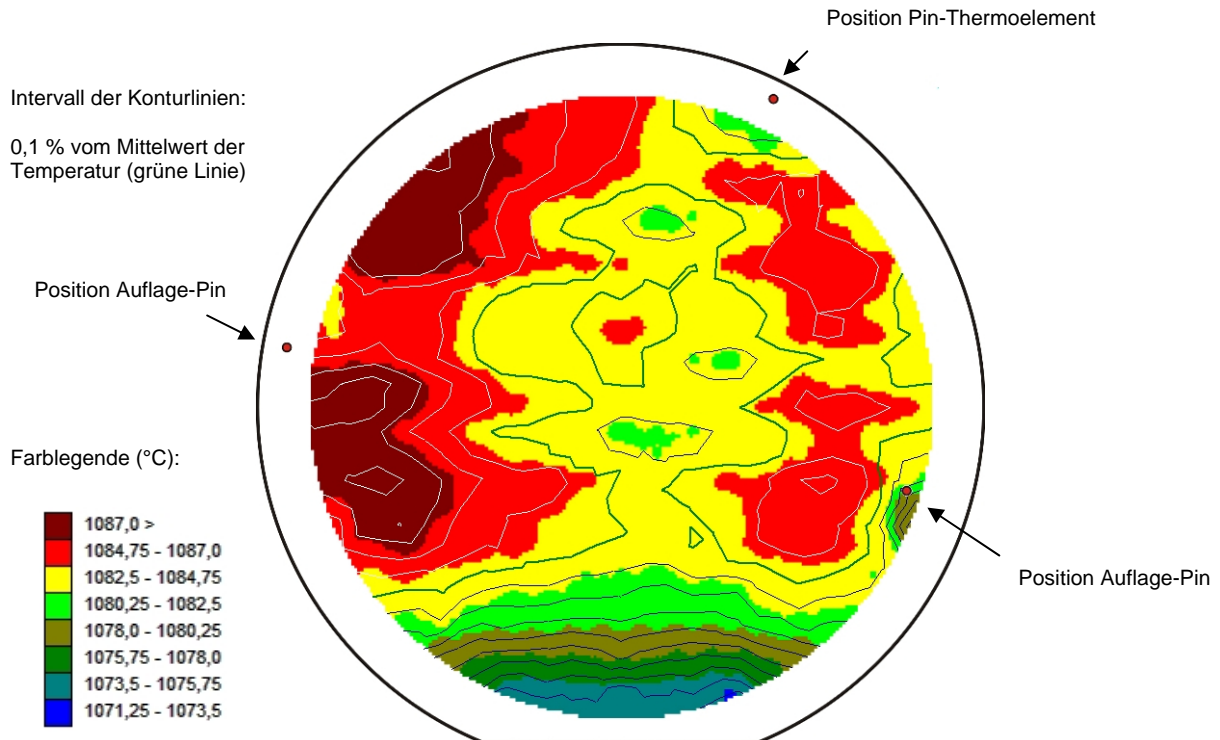


Abbildung 5.22: Darstellung der aus den Oxidschichtdickenverteilungen ermittelten lokalen Wafertemperaturen während eines statischen Oxidationsprozesses unter 40 mbar Sauerstoff (Ansicht der Waferoberseite). Die Farben entsprechen den in der Legende angegebenen Temperaturen. Eingezeichnet sind weiterhin Linien von Isothermen, wobei die grüne Linie den Temperaturmittelwert darstellt und der Abstand der dünnen Linien ein Temperaturintervall von 0,1 % angibt. Die Lage der Waferauflagepunkte ist durch rote Kreise markiert.

Die Darstellung der Temperaturverteilung in einer Höhengitterdarstellung in Abbildung 5.23 zeigt diesen Einfluss der Auflagepunkte deutlich als tiefere Bereiche nahe dem Rand.

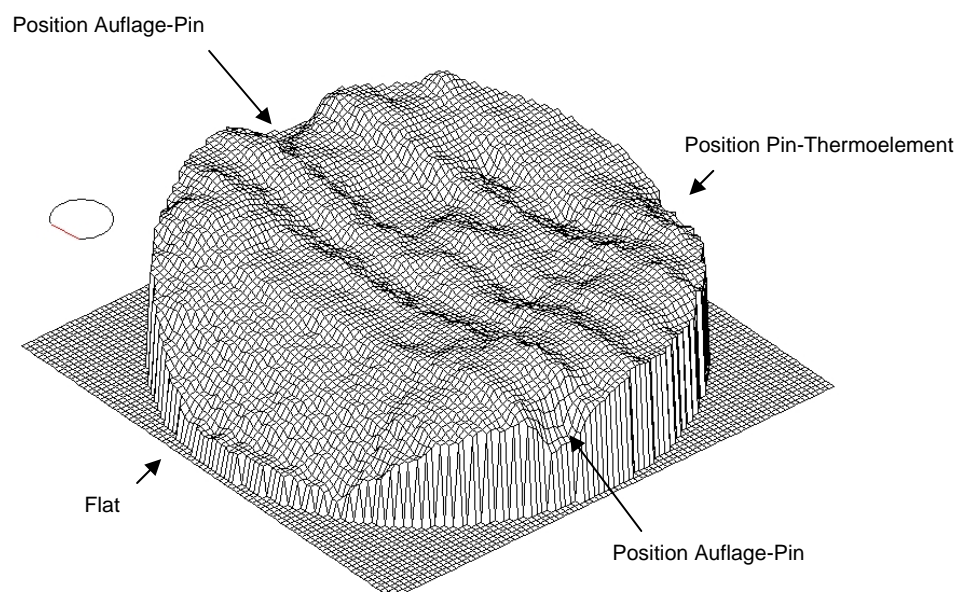


Abbildung 5.23: Wafertemperaturdarstellung im Höhengitternetz (je höher der Netzpunkt, je höher die Temperatur) des Wafers aus Abbildung 5.22. Ansicht der Waferoberseite. Eingezeichnet ist der deutliche Einfluss der Auflagepunkte auf die Temperatur.

Außerdem ist in dieser Darstellung die wellenförmige Struktur der Temperaturverteilung zu erkennen. Offenbar sind diese Schwankungen nicht zufällig und stellen keine Messfehler dar. Sie können vielmehr mit dem Einfluss der Lampen selbst begründet werden. Der Vergleich der Lampenabstände mit den Abständen der Maxima zeigt dies, wie in der Abbildung 5.24 für diesen Wafer dargestellt ist. Abgesehen von anderen lokalen Abweichungen stimmen diese Abstände überein. Dies wurde auch für andere Wafer bestätigt. Das heißt, dass die Glühwendel der Lampen als Zentren der Strahlung gewissermaßen auf den Wafer abgebildet werden und es aufgrund von dortigen lokalen Temperaturerhöhungen von ca. 2 K zu verstärktem Oxidwachstum kommt.

Offenbar ist dieser gewählte Monitoringprozess mit der oben angegebenen Abhängigkeit der Dicke von der Temperatur in dem betrachteten Temperaturbereich empfindlich genug, solche Temperaturunterschiede abzubilden.

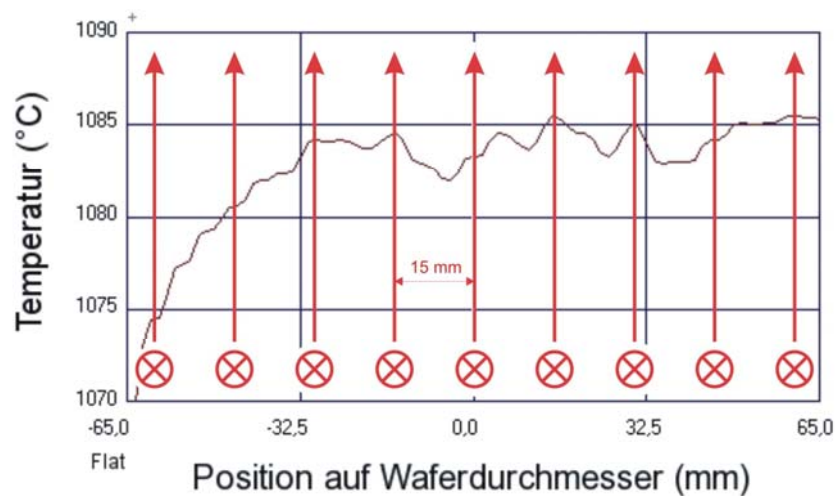


Abbildung 5.24: Bezug der Position der Lampen zu den Temperaturschwankungen. Die Stablampen sind symbolisch im Schnitt dargestellt.

Variation der Prozessparameter

Die in der Abbildung 5.20 dargestellten Daten beziehen sich auf gleiche Prozessbedingungen, mit Ausnahme der Temperaturvariation. Es sind die Ergebnisse der statischen Oxidationen bei 40 mbar ohne Leistungskorrektur gezeigt. Nicht mit dargestellt sind, der besseren Übersichtlichkeit wegen, die Werte der stationären und statischen Oxidation bei 5 mbar und die der korrigierten Lampenleistungen. Zusammengefasst ergeben sich aus diesen Daten jedoch folgende wichtige Schlussfolgerungen.

Die Oxidation bei geringerem Druck (statisch oder stationär) liefert eine geringere Oxiddicke bei gleicher Temperatur. Die Veränderung der Temperatur in dem dargestellten Temperaturbereich bewirkt eine nur geringe Änderung der Schichtdicke.

Die Schichtdicken bei den Versuchen mit Lampenkorrektur, zeigen den gleichen Zusammenhang zur Wafertemperatur, wie in der Abbildung 5.20 dargestellt. Jedoch ist die Abweichung zur geforderten (und am Pin-Thermoelement gemessene) Solltemperatur deutlich von der Korrektur abhängig. Aus der Abbildung 5.17 ist für die dort angegebene Korrektur ersichtlich, dass bei gleicher Temperaturanzeige des Pin-Thermoelements der Bereich mit der Temperaturmessstelle relativ weniger Strahlungsleistung empfangen haben muss als ohne

Korrektur. Aufgrund der Bestimmung der Isttemperatur für die Leistungsregelung an dieser Stelle hatte dies zu einer deutlichen Anhebung der Gesamtleistung geführt. Bei *gleicher* Solltemperatur wurde also mit der Korrektur der Lampenleistungen eine *höhere* Schichtdicke erzielt. Nach einer Korrektur der Lampenleistung ist somit immer auch eine Kontrolle und Kalibrierung der Temperaturmessung bei Verwendung eines Pin-Thermoelementes erforderlich.

Bei der wiederholten Ausführung gleicher Rezepte, d.h. von Prozessen mit identischer Solltemperatur, wurden teilweise unterschiedliche Schichtdickenergebnisse erzielt. Es zeigte sich nun, dass der Grund für diese Unterschiede die tatsächlich verschiedenen Wafertemperaturen waren. Es konnte aus den aufgewendeten Lampenleistungen auf die wahren Wafertemperaturen geschlossen werden, da diese in dem nach Abbildung 5.20 dargestellten erwarteten Zusammenhang zur erzielten Schichtdicke standen. So sind in dieser Abbildung drei Oxidationsexperimente mit gleicher Solltemperatur von 900 °C dargestellt, bei denen jedoch unterschiedliche Leistungen und damit Wafertemperaturen auftraten. Da es sich um gleiche Wafer mit gleichen Oberflächen handelte, kann ein Oberflächeneinfluss aufgrund unterschiedlicher Strahlungsabsorption ausgeschlossen werden. Die entsprechend erzielten Schichtdicken folgen dem parabolischen Zusammenhang. Offenbar sind bei der Randlage der Temperaturmessung bereits geringe Veränderungen in der Waferlage kritisch für das Ausführen reproduzierbarer Prozesse über eine Temperaturregelung mittels des Pin-Thermoelements.

Abweichungen von dem parabolischen Zusammenhang wurden nach Ausführen von *ersten* Oxidationsprozessen nach mehrtägigem Stillstand der Anlage, verbunden mit einem geringen Druckanstieg durch Kammerlecks, beobachtet. Bei solchen Prozessen wurden in der Regel höhere Schichtdicken erzielt als durch die aufgewendeten Lampenleistungen zu erwarten gewesen wären. Wahrscheinlich kommt es hierbei durch Spuren von Wasseradsorbaten auf den Kammerwänden, die bei anfänglichen Hochtemperaturprozessen von dort desorbieren, zu einer beschleunigten Oxidation. Bei anschließenden Prozessen trat dieser Effekt nicht mehr in Erscheinung.

5.2.4 Schlussfolgerungen

Zusammenfassend kann festgestellt werden:

- Die vorhandene Anlage weist Mängel bei der Temperaturbestimmung auf, die kritisch für eine Durchführung von reproduzierbaren Prozessen sind.
- Die Wafertemperaturen während der Prozesse weichen deutlich von den gewählten Solltemperaturen ab.
- Die Ursache dafür sind die thermische Trägheit des verwendeten Pin-Thermoelements und eine ungünstige Messposition nahe dem Waferrand.
- Die Temperaturen und die aufgezeichneten Lampenleistungen stehen jedoch in eindeutigem Zusammenhang.
- Eine Ermittlung der Wafertemperatur ist damit über die Auswertung der Lampenleistung möglich.
- Eine exakte Prozesskontrolle über die alleinige Steuerung der Lampenleistung ist jedoch kaum möglich. Die für eine genaue Temperatureinstellung notwendige Leistung ist auch abhängig von Prozessbedingungen wie Gasströmung und -art, Kontaminationen und Lampenalterung. Da dann jedoch keine Temperaturmessung über direkten Waferkontakt erforderlich wäre, entfielen der Einfluss der Waferlage (bezüglich der Messposition am Rande) auf die Leistungsanforderung.

- Die Homogenität der Temperaturverteilung ist gering und durch die Ausführung der Lampenheizung kaum optimierbar. Es ist nur ein Lampenfeld vorhanden und kein Guardring einsetzbar.
- Ein Monitoring der Temperaturverteilung ist über einen Oxidationsprozess bei einem statischen Sauerstoffdruck von 40 mbar möglich.
- Ein Oxidationsprozess bei einem Sauerstofffluss von 100 sccm und 5 mbar Prozessdruck ist für ein Monitoringprozess nicht geeignet, liefert dafür eine homogene Schichtdickenverteilung.

Damit ergeben sich auch wichtige Schlussfolgerungen für die Evaluierung der Dotierprozesse. Eine Untersuchung der Durchführbarkeit entsprechender Dotierprozesse ist möglich und die Ergebnisse sind reproduzierbar. Der Einfluss der Mängel der Temperaturmessung und -verteilung auf die Ergebnisse der Dotierprozesse ist jedoch zu berücksichtigen. Insbesondere wurden die wahren Prozesstemperaturen der Dotierversuche, wie oben beschrieben, anhand der jeweils aufgewendeten Lampenleistungen korrigiert. Die nachfolgenden Prozesstemperaturangaben beziehen sich damit auf die Temperatur der Wafermitte, wobei in der Regel die Maximaltemperatur kurz nach der Aufheizrampe als die das Prozessergebnis hauptsächlich beeinflussende Größe betrachtet wird.

5.3 Einstufige Gasphasendotierung

5.3.1 Dotiergas hoher Konzentration

Der einstufige Dotierprozess ist in der Literatur beschrieben, Sato et al. [Sato '98] berichten über Experimente mit Phosphin in Wasserstoff. Dort wurde gefunden, dass die Übergangstiefe linear mit dem Quadrat der Dotierzeit steigt, bei 1100 °C wurde bei der kürzesten untersuchten Dotierzeit von 3 min eine Übergangstiefe von 260 nm erzielt, was für die angestrebten flachen Übergänge einen zu hohen Wert darstellt. Die Oberflächenkonzentration C_{OF} von eindiffundiertem Phosphor (d.h. die Volumenkonzentration von Phosphor im Festkörper im Bereich der Oberfläche), war hingegen für die untersuchten Prozesse mit $t > 3$ min unabhängig von der Dotierzeit. Mit steigender Dotiertemperatur bis 1000 °C erhöhte sich die Oberflächenkonzentration, oberhalb von 1000 °C blieb sie jedoch konstant und war bei allen untersuchten Temperaturen auch immer weit geringer als die Festkörperlöslichkeit von Phosphor in Silicium bei der jeweiligen Temperatur (siehe Abbildung 3.1). Die Oberflächenkonzentration konnte hingegen durch Erhöhen des PH_3 -Partialdruckes gesteigert werden. Dabei war sie nur abhängig vom Phosphinpartialdruck, nicht aber vom Partialdruck des Wasserstoffs oder vom Totaldruck während des Prozesses. Der gesamte eindiffundierte Phosphor erwies sich als aktiv.

Aus diesen Ergebnissen wurde ein Modell der Wechselwirkungen von Gasphase, Abscheidung und Diffusion entwickelt. Dabei wurde angenommen, dass Phosphor aus der Gasphase in das Substrat über das Ausbilden einer Phosphor-Adsorbatschicht eindringt. Diese Adsorbatschicht wurde nachgewiesen als Zwischenschicht nach einer in-situ-Abscheidung von undotiertem Polysilicium. XPS-Messungen zeigten, dass dieser Phosphor nicht oxidiert in metallischer Form vorlag und durch HF-Behandlung entfernt werden konnte. Somit konnte vermutet werden, dass sich bei den betrachteten Hochtemperaturexperimenten ein Gleichgewicht zwischen der Gasphase und der Adsorbatschicht und der Adsorbatschicht und dem Substrat einstellte und damit die Oberflächenkonzentration ausschließlich durch den Phosphinpartialdruck bestimmt war.

Daraus wurde abgeleitet, dass für eine höhere Phosphor-Oberflächenkonzentration, wie sie für flache Übergänge mit geringem Flächenwiderstand gefordert sind, höhere Partialdrücke von Phosphin als die untersuchten von bis zu ca. 1 mbar gefordert sind.

Auch Kiyota et al. [Kiyota '94] erreichten für ihren Dotierprozess unter Atmosphärendruck (1 % PH₃ in H₂ eingemischt in einen H₂-Transportgasstrom) nur eine maximale Phosphorkonzentration von $1,3 \cdot 10^{19} \text{ cm}^{-3}$. Deren maximaler Phosphinpartialdruck war begrenzt auf ungefähr 0,06 mbar. Auch sie haben wegen der geringeren Dotiereffizienz des Phosphor-VPD im Vergleich zu ihrer entsprechend ausgeführten Bordotierung mit Diboran höhere Phosphinkonzentrationen vorgeschlagen.

So wurden zunächst mehrere Versuchsreihen mit dem höher mit PH₃-Dotiergas angereicherten Gasgemisch (ca. 10% PH₃ in N₂) durchgeführt und die Flächenwiderstände der dotierten Wafer nach den Versuchen gemessen. Die bei diesen Prozessen erreichten Phosphinpartialdrücke waren im Bereich von 0,5 bis 1 mbar. Dabei zeigte sich jedoch, dass Versuche, die mit demselben Rezept nacheinander erfolgten, nicht zu reproduzierbaren Prozessergebnissen führten, wie in der Abbildung 5.25 für vier unter gleichen Bedingungen hintereinander ausgeführte Prozesse angegeben ist.

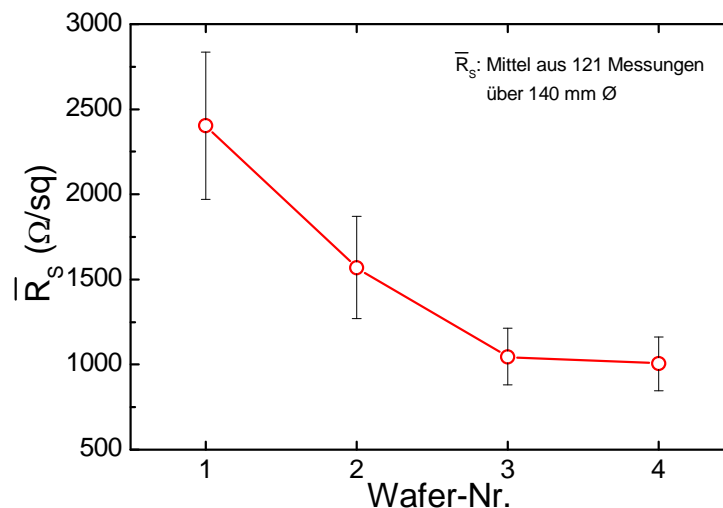


Abbildung 5.25. Mittlerer Flächenwiderstand von vier fortlaufend unter gleichen Bedingungen prozessierten Wafern ($T_{\text{Soll}} = 850 \text{ °C}$, Rampe 60/30/60 s und $p = 9 \text{ mbar}$).

Als gravierender Nachteil wurde jedoch gefunden, dass sich bereits nach nur zehn Dotierversuchen mit dem phosphinreichen Gasgemisch ein sichtbarer roter bis gelblich-roter Belag innerhalb des Reaktors bildete, sowohl an den gekühlten Wänden als auch auf der Quarzplatte. Bei diesem Belag handelte es sich mit großer Wahrscheinlichkeit um roten Phosphor, gebildet bei der thermischen Zerlegung von Phosphin.

In der Literatur sind für entsprechende Dotierexperimente nur wenige Angaben zu finden. Hsueh [Hsueh '70] berichtete von einem Auftreten von Kontaminationen mit rotem Phosphor in einem konventionellen offenen Rohrdiffusionssystem, in dem Phosphin mit einem Partialdruck von 0,3 mbar im Stickstoff-Trägergasstrom verwendet wurde. Dort wurde für die Beschreibung des Eindiffusionsvorganges von der Ausbildung einer lokalen P₂O₅-Quelle ausgegangen. Kiyota et al. [Kiyota '94] erwähnen zwar keine Kontaminationen, jedoch finden sich bei Sato et al. [Sato '98] auch bei der Verwendung des Wasserstoff-Trägergases Hinweise auf eine Kontamination der Kammerwände.

Solche Verunreinigungen mit Phosphor können jedoch gerade in RTP-Anlagen mit Lampenheizung nicht toleriert werden. Ein Belag auf dem Quarz-Fenster beeinflusst nachteilig dessen Transmissionseigenschaften, was zu einer schlechten Verteilung und Reproduzierbarkeit der in den Prozessraum eingebrachten Strahlungsleistung führt. Eine häufige Reinigung, z.B. mit einem thermischen Prozess unter Verwendung von Salzsäure, wie in [Sato '98] vorgeschlagen, würde die Anlagenkomplexität und -kosten erhöhen und wäre wohl auch nachteilig für den Dotierprozess selbst.

An dieser Stelle sind Betrachtungen zum Zerfall des Phosphins sinnvoll, um die Einflüsse abzuschätzen, die ihn bewirken. Aufgrund einer verstärkten Bedeutung des thermischen Zerfalls von Phosphin für die Erzeugung von hochreinem Phosphor für Schichtabscheidungen vor allem in Epitaxieverfahren, sind in letzter Zeit vielfältige Untersuchungen dazu veröffentlicht worden. Eine umfangreiche Zusammenfassung von neueren wichtigen Daten und Literaturangaben sind z.B. in einem Ergänzungsband zu Gmelins Handbuch der anorganischen Chemie [Behrendt '93] zusammengetragen.

Aus thermochemischen Daten wurden Gleichgewichtskonstanten für homogene Gasphasengleichgewichte bestimmt. Von Chow und Chai [Chow '83] wurde die thermische Zersetzung von reinem PH_3 in einer speziellen Hochvakuum-Pyrolyse-Apparatur für den Einsatz als hochreine Phosphorquelle beschrieben. Dafür wäre eine Generierung von dimerem P_2 anstelle von tetramerem P_4 von Vorteil, da davon auszugehen ist, dass P_2 an kalten Oberflächen eher in der Form von ungiftigem roten Phosphor kondensiert als in der hochgiftigen weißen Form [Wright '82]. Dies ist natürlich für den Einsatz in technischen Prozessen von Vorteil. Außerdem besitzt roter Phosphor einen bedeutend niedrigeren Dampfdruck von ca. $2,5 \cdot 10^{-10}$ mbar (bei Raumtemperatur) als der bereits bei ca. 44°C schmelzende weiße Phosphor mit einem Dampfdruck von $5 \cdot 10^{-2}$ mbar (bei Raumtemperatur), so dass Wandkontaminationen aus rotem Phosphor zwar weniger flüchtig und damit kaum über moderates Ausheizen zu entfernen sind, jedoch somit auch kaum eine Quelle von Querkontaminationen in nachfolgenden Prozessen sind. Für den Zerfall in der Gasphase wurden folgende Zerfallsreaktionen betrachtet.



Aus dem Vergleich der aus thermodynamischen Daten bestimmten Gleichgewichtskonstanten der jeweiligen Reaktionen wurde erwartet, dass bis zu einer Temperatur von 1473°C der Zerfall zu P_4 überwiegt, oberhalb davon jedoch der zu P_2 . Die Bildung von atomarem Phosphor wäre hingegen unwahrscheinlich, selbst bei 1700°C überwiegen die molekularen Formen um mehr als das 1000fache. Deren Gleichgewichtskonstanten sind für Temperaturen oberhalb von 600°C selbst bei sehr geringen Phosphorpartialdrücken größer als 10, so dass ein hoher Anteil von Phosphin bei diesen Temperaturen bereits in der Gasphase zerfallen sein sollte. Aus dem Zerfallsgleichgewicht von P_4 ,



das selbst druckabhängig ist, wobei mit geringeren Phosphorpartialdrücken der Zerfall begünstigt ist, wurde gefolgert, dass selbst bei geringen Phosphorpartialdrücken von 1 mbar ab 800°C dieser Zerfall hierbei zum günstigeren P_2 überwiegt.

Aus einer eingehenderen thermodynamischen Betrachtung der Zerfallsgleichgewichtsreaktion (62) kann nun abgeschätzt werden, welcher Anteil von Phosphin (unter der Annahme des Vorliegens von Gleichgewichtsbedingungen) bereits in der Gasphase zerfallen sein sollte. Korrigierte Werte der NIST-JANAF Thermochemischen Tabellen [Chase '98] für diese Reaktion wurden von Lodders [Lodders '99] angegeben. Aus den Angaben der Bildungsreaktion von $\text{PH}_3(\text{g})$, die der Rückreaktion von Gleichung (62) entspricht, ergeben sich die in der Abbildung 5.26 dargestellten Werte der Gleichgewichtskonstanten K_p ,

$$K_p = \frac{p_{\text{P}_2}^{1/2} p_{\text{H}_2}^{3/2}}{p_{\text{PH}_3}}, \quad (65)$$

in Abhängigkeit von der Temperatur für diese Reaktion.

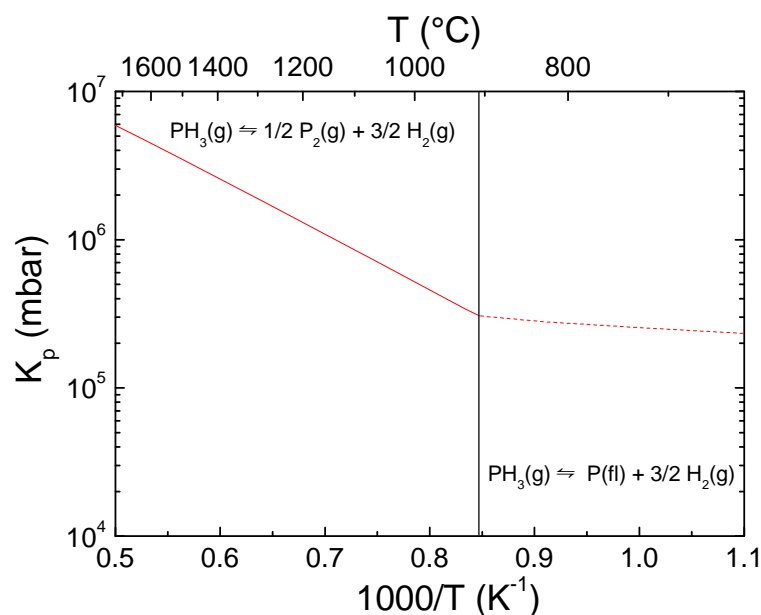


Abbildung 5.26: Darstellung der Gleichgewichtskonstanten K_p der Zerfallsreaktion von $\text{PH}_3(\text{g})$ in Abhängigkeit von der reziproken Temperatur. Berechnet aus Daten von [Lodders '99] für 1000 mbar, oberhalb von 1180 K wurde dort der Phosphorreferenzzustand $\text{P}_2(\text{g})$ wie bei der Betrachtung von Gleichung (62) benutzt.

Aus dem Verlauf von K_p (im Folgenden soll nur der in der Abbildung links dargestellte Bereich oberhalb der Verdampfungstemperatur von flüssigem Phosphor bei 1 bar betrachtet werden) ergibt sich, dass mit steigender Temperatur das Gleichgewicht zum Zerfall von PH_3 hin verschoben wird. Aus den großen Werten von K_p für den gesamten betrachteten Temperaturbereich folgt jedoch, dass bei Normaldruck der Zerfall von PH_3 praktisch immer vollständig wäre. Bei Einsatz von Phosphin in einer Wasserstoffatmosphäre wird aber das Gleichgewicht entgegen einem Zerfall verschoben. Aus der Betrachtung der Volumenänderungen der Reaktion folgt hingegen, dass mit geringerem Druck wiederum der Zerfall von PH_3 günstiger wird. Da für Gasphasendotierprozesse Phosphin sowohl bei sehr geringen Drücken als auch unter Verwendung von Wasserstoffträgergas verwendet wird (siehe folgendes Kapitel), lohnt es sich trotz des sehr hohen Wertes von K_p eine Einschätzung der Auswirkungen typischer Prozessbedingungen auf den Zerfall von Phosphin in der Gasphase vorzunehmen.

Dafür soll im Folgenden als Beispiel eine Prozesstemperatur von 1000 °C betrachtet werden. Für diese Temperatur ergibt sich für die Zerfallsreaktion (62) aus Abbildung 5.26 ein K_p -Wert

von 520 575 mbar. Bei Einsatz von Phosphin mit einem Partialdruck von p_{PH_3} und Wasserstoff mit einem Partialdruck von p_{H_2} folgt aus dem Massenwirkungsgesetz bei einem Stoffumsatz x (hier ausgedrückt durch den entsprechenden Partialdruckumsatz) gemäß der Reaktion (62) bis zum Erreichen des Gleichgewichts:

$$520\,575 \text{ mbar} = \frac{(0,5 \cdot x \text{ mbar})^{\frac{1}{2}} (p_{H_2} + 1,5 \cdot x \text{ mbar})^{\frac{3}{2}}}{(p_{PH_3} + x \text{ mbar})} \quad (66)$$

Durch Auflösen dieser Gleichung nach x kann für verschiedene anfängliche Partialdrücke deren Einfluss auf den Gleichgewichtspartialdruck von Phosphin, bzw. den Anteil des zerfallenen Phosphins bestimmt werden:

$$PH_{3(\text{zerfallen})} \% = \frac{x_{\text{Umsatz } PH_3} \text{ mbar}}{p_{PH_3}} \cdot 100. \quad (67)$$

In der Abbildung 5.27 ist für drei verschiedene Gesamtdrücke $p_{PH_3} + p_{H_2}$ die Abhängigkeit des Phosphinzerfalls vom eingesetzten Phosphinpartialdruck in Wasserstoff dargestellt. Der sich bei der Verwendung einer geringen Konzentration vom 50 vpm Phosphin in Wasserstoff bei den entsprechenden Gesamtdrücken ergebende anfängliche Phosphinpartialdruck ist in der Abbildung angegeben.

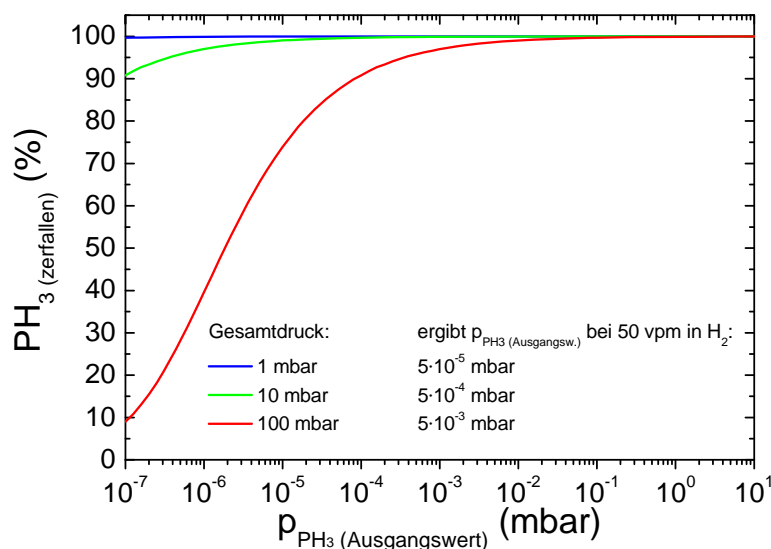


Abbildung 5.27: Aus der Gleichgewichtsbetrachtung der homogenen Gasphasenreaktion (62) ermittelter Anteil von zerfallenem Phosphin bei einer Temperatur von 1000 °C in Abhängigkeit vom eingesetzten Phosphinpartialdruck in Wasserstoffatmosphäre bei drei verschiedenen Gesamtdrücken. Angegeben ist der bei Verwendung von 50 vpm PH_3 in H_2 bei dem jeweiligen Gesamtdruck vorherrschende anfängliche Phosphinpartialdruck.

Hier zeigt sich nun, dass selbst unter Verwendung des verdünnten Phosphins in Wasserstoff bei den angegebenen Prozessdrücken dieses (bei der Möglichkeit der Einstellung des chemischen Gleichgewichts in der Gasphase) bei einer Temperatur von 1000 °C praktisch vollständig in gasförmiges P_2 und Wasserstoff zerfallen sein sollte. Dies gilt dann natürlich erst recht bei der Verwendung von reinem Phosphin oder auch Phosphin in Stickstoff bei den angegebenen Partialdrücken.

Hierbei ist jedoch zu beachten, dass dies Gleichgewichtsbetrachtungen bei homogenen Reaktionen in der Gasphase sind. Diese Gleichgewichtsbedingungen werden bei den realen

Systemen jedoch nicht immer vorliegen. So kann aus den thermodynamischen Daten abgeleitet werden, dass Phosphin selbst bei Raumtemperatur in festen weißen Phosphor P_4 zerfallen sollte. Dies wird experimentell jedoch nicht beobachtet, aufgrund einer erforderlichen Aktivierungsenergie für die Zerfallsreaktion.

Doch selbst bei den oben betrachteten hohen Prozesstemperaturen ergeben sich bei realen Experimenten erhebliche Abweichungen von den theoretischen Annahmen. So fanden Chow und Chai [Chow '83] bei ihren Versuchen im Temperaturbereich von 100 - 1070 °C je nach eingesetztem Heizerflächenmaterial nur eine Zerfallseffizienz¹ des Phosphins von 77 - 90 % bezogen auf den Phosphinpartialdruck bei Raumtemperatur, obwohl ab ca. 300 °C das Phosphin bereits komplett zerfallen sein sollte. Bei Heizflächen aus Tantal zeigte sich ein Peak in der Phosphorkonzentration im Bereich um 400 °C, bei Quarzflächen hingegen erfolgte ein deutlicher Zerfall von PH_3 ab ca. 550 °C, kontinuierlich steigend bis zu einer Sättigung bei ca. 1000 °C. Als vorherrschende Phosphorphase nach PH_3 -Zerfall wurde P_2 gefunden. Aus den Abweichungen von den theoretischen Betrachtungen wurde gefolgert, dass keine Gleichgewichtsbedingungen in einer homogenen Gasphase vorlagen, insbesondere wurde keine Abhängigkeit der Zerfallsanteile von den Partialdrücken im untersuchten Bereich gefunden.

So zeigt sich hier und aus weiteren experimentellen Untersuchungen, dass der Phosphinzerfall sehr komplexe Vorgänge beinhaltet [Behrendt '93]. Der Zerfall ist kinetisch gehemmt, eine erhöhte thermische Zersetzung wurde im Allgemeinen erst oberhalb von 400 °C beobachtet. Als Zerfallsprodukte sind H_2 , P_2 und P_4 aber auch Zwischenprodukte wie PH_2 und PH in der Gasphase gefunden worden. Offenbar handelt es sich bei der thermischen Zersetzung in den meisten Fällen um eine heterogene, oberflächenkatalytische Reaktion. Daraus ergibt sich insbesondere auch ein messtechnisches Problem. Ex-situ Analysetechniken der Gaszusammensetzung, wie sie häufig mittels Gasanalyse-Massenspektrometern in Entfernung von heißen Oberflächen ausgeführt werden, besitzen vielfach selbst einen Einfluss auf die Ergebnisse. Bei Einleitung von Phosphin im Wasserstoff-Trägergas durch eine Heizzone, wie sie u.a. bei Epitaxiesystemen üblich ist, zeigten sich deutliche Unterschiede im Zerfallsgrad von PH_3 zwischen in-situ Analysen in diesem Bereich, wie z.B. durch laserinduziertes Raman-Scattering, und den ex-situ Analysen von zum Massenspektrometer gepumpten Prozessgas. Ex-situ Messungen ergeben dann, dass eine vermeintlich höhere Temperatur für den Zerfall erforderlich sein müsste, als es die in-situ Ergebnisse zeigen. Das heißt, dass heiße Oberflächen der Katalysator für den Zerfall sind und es im Gasstrom zur ex-situ Analyse bereits wieder zu Bildungsreaktionen kommt. Da P_4 bei solchen Untersuchungen thermodynamisch stabiler als PH_3 ist, kann eine Bildung nur aus PH_2 oder PH erfolgen. Auch zeigte die Analyse des H_2 -Ertrags, dass PH_3 nicht vollständig zerfallen sein konnte. Außerdem zeigt sich, dass nicht nur Temperatur, Druck und Zusammensetzung der Gasmischung Einfluss auf den Zerfall haben, sondern dieser auch von Material und Oberfläche des Reaktors, seiner Geometrie und Vorgeschichte, von Flussraten durch den und Aufenthaltszeiten im heißen Bereich des Reaktors abhängig ist. Die Art des Trägergases hat dabei keinen Einfluss auf die Zerfallstemperatur.

Der Mechanismus der oberflächenkatalytischen Reaktion ist noch nicht genau geklärt. Deren Aktivierungsenergie variiert stark in Abhängigkeit vom Oberflächenmaterial und Temperaturbereich. Verminderte Aktivierungsenergien durch Zerfallsmechanismen, die eine dissoziative Adsorption an Oberflächen einschließen, wurden beobachtet (siehe auch dazu Kapitel 3.1.3). Bei der Verwendung von Lampenheizungen ist prinzipiell auch ein Zerfall durch UV-Photolyse möglich. Dieser sollte für Wellenlängen unterhalb von 230 nm stattfinden. Für

¹ Die Untersuchungen zur Phosphinpyrolyse für den Einsatz als hochreine Phosphorquelle hatten eine hohe Effizienz des Phosphinzerfalls in der Gasphase zum Ziel, während für die Dotierexperimente zur Vermeidung von Kontaminationen die Stabilität des Phosphins in der Gasphase wichtig ist.

solche Zersetzungen wurden auch feste Abscheidungen rötlicher bis gelblicher Farbe an Reaktorwänden gefunden, qualitative Analysen zeigten darin eine Mischung aus Phosphor-Allotropen und Polyphosphanen an. Die Energieabstrahlung der verwendeten Infrarot-Halogenlampen ist in diesem Wellenlängenbereich jedoch vernachlässigbar gering.

5.3.2 Dotiergas geringer Konzentration

Es ergab sich also bereits nach den anfänglichen Dotierexperimenten, dass der einstufige VPD-Prozess unter Verwendung eines phosphinreichen Dotiergases technologisch ungünstig ist. Dieser Ansatz wurde daraufhin nicht weiter verfolgt.

Es zeigte sich jedoch, dass eine hohe Phosphinkonzentration in der Gasphase für einen VPD-Prozess erforderlich ist, um ausreichende Phosphorkonzentrationen im Silicium zu erhalten. Die Verwendung von 50 vpm Phosphin in Wasserstoff führte bei gleichem Prozessablauf nicht zu einem positiven Dotierergebnis, wie Flächenwiderstandsmessungen zeigten. Dies war auch der Fall, wenn der Druck des Dotiergasgemisches in der Kammer auf 100 mbar erhöht wurde, was einem Phosphinpartialdruck von ca. $5 \cdot 10^{-3}$ mbar entspricht. Das bedeutet, dass diese geringen Phosphinkonzentrationen für einen solchen Dotierprozess nicht ausreichend sind.

Ein Vergleich mit den Ergebnissen von Sato et. al. [Sato '98], wie sie auch für die Diffusion in Wasserstoffatmosphäre gefunden wurden, zeigt den zugänglichen Partialdruckbereich bei den Dotierexperimenten mit der geringen Phosphinkonzentration. In der Abbildung 5.28 ist der von Sato et. al. gefundene Zusammenhang in dem von ihnen untersuchten Bereich angegeben und zu geringeren Partialdrücken hin extrapoliert worden.

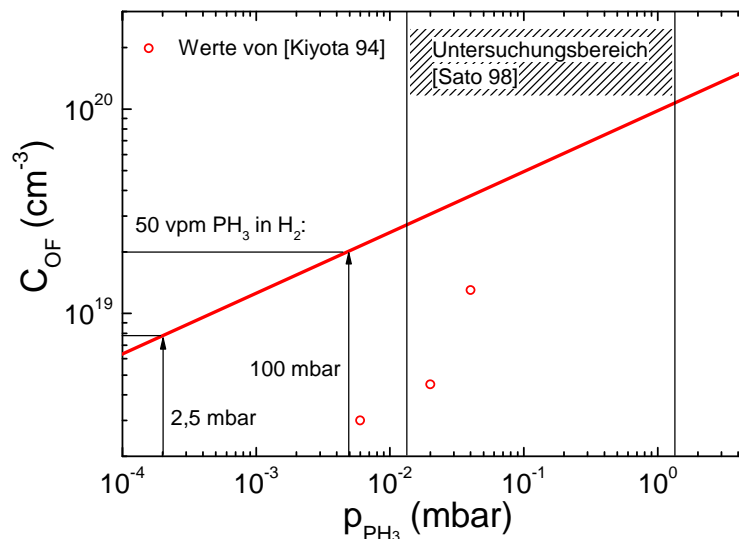


Abbildung 5.28: Abhängigkeit der Oberflächenkonzentration von Phosphor in Silicium nach [Sato '98] für eine Gasphasendiffusion von Phosphin in Wasserstoff (1000 °C, 3 min) extrapoliert auf die Partialdruckbereiche der durchgeführten Experimente mit 50 vpm PH_3 in H_2 . Zum Vergleich sind Ergebnisse von [Kiyota '94] für eine Gasphasendiffusion unter Atmosphärendruck bei 900 °C und 30 min mit angegeben.

Hier zeigt sich, dass, unter der Annahme der Gültigkeit des von Sato et al. gefundenen Zusammenhangs auch für niedrigere Partialdrücke als die dort untersuchten, selbst die maximalen Phosphor-Oberflächenkonzentrationen, die ab Temperaturen oberhalb von

1000 °C erzielt wurden, bei der Verwendung des geringer konzentrierten Phosphingasgemisches zu ineffizient für eine effektive Dotierung sind. Auch der in der Abbildung gezeigte Vergleich einer Gasphasendotierung von Kiyota et al. für 30 min unter Atmosphärendruck bei 900 °C zeigt die geringe Effizienz solcher Methode, die zu einer Phosphorkonzentration weit unterhalb der Festkörperlöslichkeit von Phosphor in Silicium (bei der Prozesstemperatur) führt.

Im Kapitel 3.2.2 wurde gezeigt, dass die Belegung mit nur einer Monolage Phosphor bei einer idealen Dotierung eine ausreichende Dosis für die angestrebten Übergänge darstellte. Für einen Wafer von 150 mm Durchmesser sind dafür ca. $1,2 \cdot 10^{17}$ Phosphoratome erforderlich. Unter Verwendung der unterschiedlichen Dotiergaskonzentrationen kann abgeschätzt werden, welche Dotierstoffmenge bei einem Kurzzeitprozess bereitgestellt wird.

Bei einem Fluss F des Dotiergases mit dem Phosphinanteil C_{PH_3} für eine Zeit t ergibt sich ein Volumen von Phosphin (unter Standardbedingungen) von

$$V_{nPH_3} = F C_{PH_3} t. \quad (68)$$

Mit dem molaren Volumen $V_m = 22,4 \frac{\ell}{\text{mol}}$ und der Avogadrozahl $N_A = 6,022 \cdot 10^{23} \text{ mol}^{-1}$ folgt aus der Stoffmenge des im Gas enthaltenen Phosphins,

$$n_{PH_3} = \frac{V_{nPH_3}}{V_m}, \quad (69)$$

die Anzahl der enthaltenen Phosphinmoleküle:

$$N_{PH_3} = n_{PH_3} N_A = \frac{F C_{PH_3} t}{V_m} N_A. \quad (70)$$

Da pro Phosphinmolekül ein Phosphoratom vorhanden ist, entspricht diese auch der Anzahl der bereitgestellten Phosphoratome.

Als Beispiel soll ein Gasfluss von 100 sccm für eine Minute angenommen werden. Damit ergeben sich für

$$\begin{array}{ll} 10 \% \text{ PH}_3 \text{ im Gasfluss } (C_{PH_3} = 0,1): & N_P = 2,7 \cdot 10^{20} \\ 50 \text{ vpm PH}_3 \text{ im Gasfluss } (C_{PH_3} = 5 \cdot 10^{-5}): & N_P = 1,3 \cdot 10^{17}. \end{array}$$

Mit dem höher konzentrierten Gasgemisch werden somit innerhalb von nur einer Minute bereits das über 2000fache der benötigten Dotierstoffmenge eingebracht, während die Menge in dem gering konzentrierten Gas ungefähr der benötigten Menge entspricht. Da aber offenbar selbst die mit dem höher konzentrierten Gasgemisch erzielten Partialdrücke nicht für einen ausreichenden Phosphoreintrag in das Silicium genügen, ist die einstufige Gasphasendotiermethode für Phosphor als äußerst ineffektiv anzusehen.

5.3.3 Schlussfolgerungen

Zusammengefasst folgt aus den Betrachtungen zur einstufigen Gasphasendotierung:

- Der Zerfall von Phosphin in der Gasphase ist bei hohen Temperaturen thermodynamisch günstig. Selbst bei Verwendung von Wasserstoffatmosphären ist dieser unter relevanten Prozessbedingungen zu erwarten.
- Bis hinauf zu ca. 550 °C ist jedoch noch ein hoher Anteil von Phosphin in der Gasphase vorhanden.
- Der Zerfall erfolgt vorrangig zu gasförmigem P₂, das unter Abscheidung eines roten Belages an kühlen Kammerwänden kondensiert. Das Kondensat ist stabil und nicht flüchtig.
- Die Kontaminationen sind technologisch ungünstig. Die Reduzierung von Phosphorkontaminationen ist nur durch eine Verminderung der absoluten Menge des eingesetzten Phosphins bei geringen Temperaturen möglich.
- Eine hohe Dotiergaskonzentration ist jedoch für eine effiziente Dotierung mit der einstufigen Gasphasenmethode erforderlich.

Wegen dieser Probleme der einstufigen Gasphasendiffusion sollte nun ein Dotierprozess gefunden werden, der auch mit geringen Phosphinkonzentrationen in einer Gasphase zu positiven Dotierergebnissen führt.

5.4 Zweistufiger Dotierprozess

5.4.1 Verfahren, Ablauf und angewendete Effekte

Aufgrund der Nachteile des einstufigen Dotierverfahrens als reiner Gasphasenprozess wurde ein alternativer Ansatz entwickelt, bei dem der Dotierprozess in zwei Stufen abläuft. Ähnliche Methoden sind in der Literatur für die Dotierung von Silicium mit Bor und Arsen beschrieben [Kim '00; Koh '02]. Dort wurde jedoch eine zusätzliche SiO₂-Deckschicht abgeschieden, um Desorption und Ausdiffusion der Dopanten zu verhindern. In der im Folgenden beschriebenen Methode wurde ein solcher Ansatz für die Phosphordotierung angewendet. Dabei wurde auch untersucht, ob ein zusätzlicher Decklayer vermieden werden kann.

Für den Zweistufen-Dotierprozess wurden zwei unterschiedliche Abläufe gewählt, die sich vorrangig in der Durchführung des Belegungsschrittes unterscheiden. Sie sollen im Folgenden vorgestellt werden. Die Abbildung 5.29 und Abbildung 5.30 zeigen jeweils den typischen Verlauf der Solltemperaturen¹, wie sie in den zwei Prozessvarianten eingestellt wurden. Für diese Versuche wurde ausschließlich das in Wasserstoff hoch verdünnte Phosphin verwendet.

¹ Für die Beschreibungen der Abläufe soll an dieser Stelle die Solltemperatur genügen, aufgrund der Probleme der Temperaturbestimmung, wie in Kapitel 5.2 dargelegt, ergaben sich jedoch höhere Prozesstemperaturen.

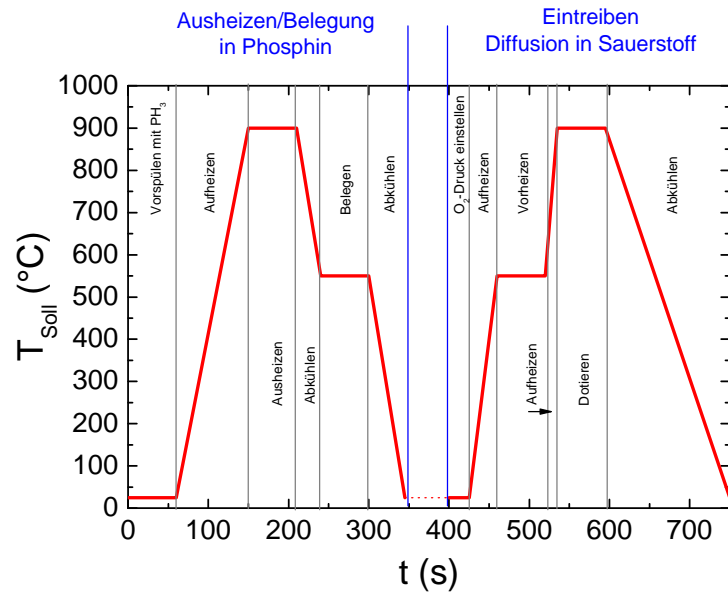


Abbildung 5.29: Typischer Ablauf des Zwei-Stufen-Dotierverfahrens der Prozessvariante 1 mit einem Ausheizschritt vor der Belegung, dargestellt anhand des Verlaufs der Solltemperatur.

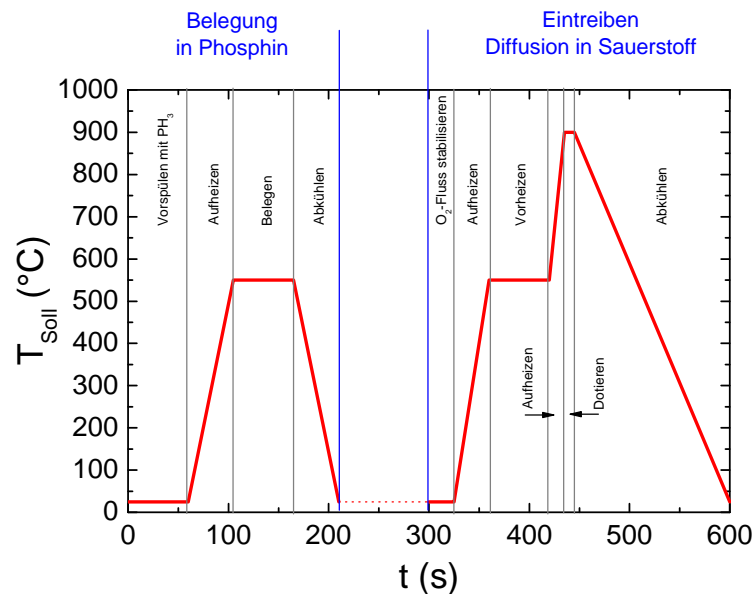


Abbildung 5.30: Typischer Ablauf des Zwei-Stufen-Dotierverfahrens der Prozessvariante 2 ohne Ausheizschritt vor der Belegung, dargestellt anhand des Verlaufs der Solltemperatur.

1. Stufe: Belegung

In Prozessvariante 1 wurde der Wafer mit einer schnellen Rampe auf 900 °C erhitzt, während der Fluss von PH_3/H_2 konstant gehalten wurde, so dass sich in der Kammer ein Gleichgewichtsdruck von 2,5 mbar einstellte. Das anfängliche Ausheizen wurde durchgeführt, um die Waferoberfläche zu konditionieren. Vom Ausheizen in Wasserstoffatmosphäre ist bekannt, dass damit eventuell eine noch vorhandene natürliche Oxidschicht entfernt werden kann [Sato '98]. In dieser Prozessvariante wurde dafür gleich die Dotieratmosphäre gewählt, da sie zum größten Teil aus Wasserstoff bestand und die vorangegangenen Versuche gezeigt hatten, dass damit noch keine nennenswerte Dotierung möglich war. Durch den Verzicht auf eine zusätzliche Gaslinie sollte die Prozessführung vereinfacht werden, und es wurde eine günstigere Dotiereffizienz erhofft.

Nach dem Ausheizschritt wurde die Wafertemperatur auf 550 °C gehalten, der Temperatur der maximalen Phosphorbedeckung an der Siliciumoberfläche. Dies war der eigentliche Belegungsschritt durch die thermische Dissoziation von PH₃ an der Siliciumoberfläche. Die Sättigung der Oberflächenbelegung mit Phosphor wird bereits mit geringen PH₃-Dosen von 12 - 16 L erreicht¹ [Yu '84; Tsukidate '99], so dass hier das hoch verdünnte Dotiergas verwendet werden kann.

In Prozessvariante 2 wurde auf den Ausheizschritt verzichtet. Hier erfolgte sogleich eine Belegung der Oberfläche bei 550 °C in der Dotiergasströmung.

2. Stufe: Eintreiben

Das Eintreiben des Phosphors in das Silicium wurde jeweils in sauerstoffhaltiger Atmosphäre durchgeführt. In dieser Prozessstufe wurde kein weiteres Dotiergas zugeführt. Der Sauerstoff sollte die Eindiffusion des Oberflächenphosphors fördern, ohne dass eine Oxiddeckschicht, wie bei CVD-Verfahren üblich, auf der Oberfläche deponiert wird. Der Sauerstoff sollte mit dem Silicium reagieren und damit den Phosphor von der Oberfläche verdrängen. An Grenzschichten von Oxid und Silicium kommt es für Phosphor infolge einer Segregation zu einer Abreicherung im Oxid und einer Anreicherung im Silicium. Dies sollte für einen Phosphoreintrag von der Oberfläche günstig sein. Außerdem beeinflusst die Anwesenheit von Sauerstoff die Diffusion von Dotierstoffen. Sauerstoff-verstärkte Diffusion (OED - oxidation enhanced diffusion) in Silicium wird für Dopanten beobachtet, die vorzugsweise über einen Interstitial-Mechanismus diffundieren, so auch für Phosphor, aufgrund der durch die Oxidation induzierten Generation von überschüssigen Eigenzwischengitteratomen im Silicium.

In Prozessvariante 1 wurde dafür Sauerstoff bei geschlossenem Kammerventil bis zu einem bestimmten Druck in die Prozesskammer eingelassen. Das Eintreiben bei der hohen Dotiertemperatur erfolgte daraufhin in einer statischen Atmosphäre.

Bei Prozessvariante 2 wurde das Eintreiben mit kontinuierlichem Sauerstofffluss durchgeführt. In beiden Fällen erfolgte zunächst ein Vorheizschritt bei 550 °C, bevor der eigentliche Dotierschritt bei der hohen Prozess Temperatur erfolgte.

5.4.2 Prozessanalyse: Dotiererfolg, Homogenität und Reproduzierbarkeit

Für eine Auswertung des Dotiererfolges und der Prozessqualität sind Flächenwiderstandsmessungen mittels der Vierspitzenmethode sehr gut geeignet. Sie sind einfach durchzuführen, ohne dass der Wafer zerstört werden muss. Messungen der Dotierhomogenität sind fast über die gesamte Waferausdehnung durch orts aufgelöste Messungen möglich. Ein messbarer Flächenwiderstand ist hierbei ein einfacher Nachweis einer erfolgten n-Dotierung im verwendeten p-Typ Grundmaterial. Ein mit dem verwendeten Vierspitzen-Messgerät nicht-messbarer Flächenwiderstand heißt damit zwar nicht unbedingt, dass keine Dotierung erfolgte, jedoch wäre diese dann so gering, dass sie nicht ausreichte, die Grunddotierung zu kompensieren. Für die beschriebenen Ziele der untersuchten Dotierverfahren, sind solche „Dotierungen“ nicht relevant. Somit soll im Folgenden zunächst der Einfachheit halber generell der „Erfolg“ der Dotierung anhand von Flächenwiderstandsmessungen bewertet werden und entsprechend deren Ergebnissen von erfolgter, hoher oder geringer Dotierung gesprochen werden.

¹ Ein verbreitetes Maß für die Dosis bei der Einwirkung von Gasatmosphären auf Oberflächen ist das Langmuir. $1 \text{ L} = 1 \cdot 10^{-6} \text{ Torr} \cdot \text{s} = 1,33 \cdot 10^{-6} \text{ mbar} \cdot \text{s}$.

Prozessvariante 1

Nach den Experimenten der Prozessvariante 1, die zuerst durchgeführt wurden, ergaben sich folgende allgemeine Ergebnisse:

- Eine Phosphordotierung des Siliciums ist unter Verwendung des gering in Wasserstoff konzentrierten Phosphins im Gegensatz zu dem einstufigen Verfahren möglich. Die erzielten Flächenwiderstände der n-dotierten Oberflächenbereiche lagen in Abhängigkeit von den Prozessparametern im Bereich von 0,9 bis 50 k Ω /sq.
- Eine sichtbare Kontamination der Reaktorinnenflächen wurde nach ca. 25 Versuchen nicht beobachtet.
- Eine Sauerstoffatmosphäre ist für das Eintreiben des Phosphors in das Silicium erforderlich. Eine gleichartige Durchführung des Eintreibeschrittes ohne die Verwendung von Sauerstoff führte zu keinem positiven Dotierresultat. Ein Sauerstoffdruck von 2,5 mbar ist hingegen für ein Eintreiben bereits ausreichend.
- Ein Abscheiden einer zusätzlichen Deckschicht zur Verhinderung der Phosphordesorption kann damit vermieden werden.
- Die Homogenität der Dotierung bezüglich der Waferfläche als auch die Reproduzierbarkeit von Prozessen, die unter gleichen Versuchsbedingungen durchgeführt wurden, erwiesen sich als sehr mangelhaft. Die Standardabweichung der Messwerte im Messbereich von ca. 130 mm Durchmesser der Waferoberfläche variierte von 9,3 % bis zu 21,7 %.

Dem erfolgreichen Nachweis der prinzipiellen Machbarkeit der Zwei-Stufen-Gasphasendotierung standen mit den im letzten Punkt beschriebenen Beobachtungen somit einige gravierende Nachteile gegenüber, die insbesondere eine weitere quantitative Auswertung der einzelnen Dotierergebnisse erschwerten. Ein eindeutiger, planbarer und nachvollziehbarer Einfluss von Veränderungen von Prozessparametern auf die Resultate war damit kaum oder nur sehr eingeschränkt zu ermitteln. Aus diesem Grund soll hier auf Darstellungen der Abhängigkeiten von durchgeführten gezielten Variationen von Prozesseinflüssen wie Prozesstemperatur, -druck und -zeit auf die erzielten Flächenwiderstände verzichtet werden. Die in Kapitel 5.2.3 aufgezeigten Mängel der Anlage waren für diese Prozessinstabilitäten von großer Bedeutung. Die folgenden Beispiele sollen dies illustrieren.

Zur Verdeutlichung der Homogenität der Flächenwiderstände sind in den folgenden Abbildungen die Verläufe der Flächenwiderstände über den Waferdurchmesser für zwei mit identischen Rezepten durchgeführte Dotierversuche dargestellt. Die Lage dieser Verläufe wurde so wie in den Oxidationskalibrierversuchen gewählt. Die Dotierversuche wurden noch ohne die Lampenkorrektur durchgeführt.

Die Verläufe zeigen eine große Inhomogenität des Flächenwiderstands bezüglich der Waferausdehnung an. Aus den Oxidationsversuchen konnte geschlossen werden, dass der Waferbereich in Richtung des Gaseinlasses/Beladeports kühler ist. Bei dem Wafer aus Abbildung 5.31 führte dies zu einem erwarteten höheren Flächenwiderstand in diesem Bereich. Die gefundene Inhomogenität ist jedoch zu groß, um allein durch die ermittelten Temperaturdifferenzen begründet werden zu können.

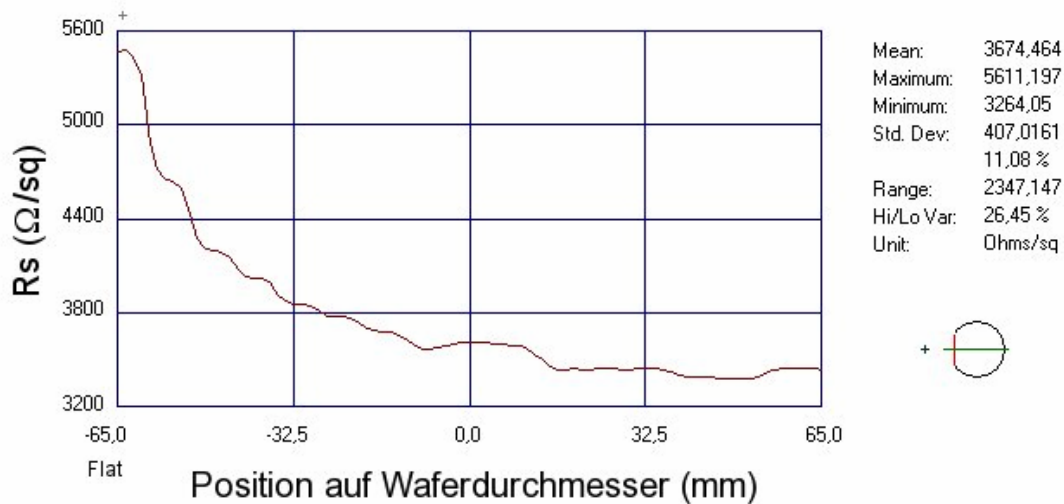


Abbildung 5.31: Darstellung des Verlaufs des Flächenwiderstands über den Waferdurchmesser (rechts unten symbolisiert), Eintreibprozess bei 1000 °C Maximaltemperatur (ermittelt aus Lampenleistung) und 60 s Dotierzeit (Wafer 3 in Abbildung 5.33). Die Lage des Verlaufs bezüglich der Lampen ist wie bei den Oxidationsversuchen. Die Angaben zur Statistik beziehen sich auf den gesamten Testdurchmesser von ca. 130 mm.

Bei einem später prozessierten Wafer ergibt sich darüber hinaus ein anderes Bild, wie in Abbildung 5.32 dargestellt ist. Bei diesem Prozess war die Temperatur etwas höher, wie aus den aufgewendeten Lampenleistungen ermittelt wurde, was zu einem geringeren mittleren Flächenwiderstand führte. Der Verlauf der Inhomogenität ist jedoch nicht der, wie er aus der Temperaturverteilung zu erwarten wäre. Offensichtlich führen noch andere Prozessparameterinhomogenitäten als die der Wafertemperatur zu unbefriedigenden Prozessergebnissen.

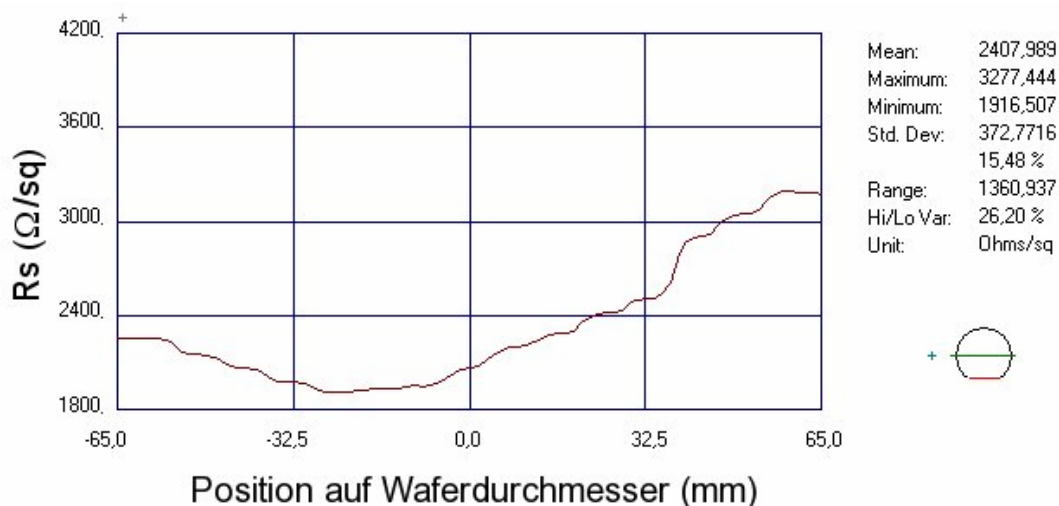


Abbildung 5.32: Darstellung des Verlaufs des Flächenwiderstands über den Waferdurchmesser (rechts unten symbolisiert), Eintreibprozess bei 1032 °C Maximaltemperatur (ermittelt aus Lampenleistung) und 60 s Dotierzeit (Wafer 9 in Abbildung 5.33). Der Flat war hier um 90° gegenüber Wafer 3 gedreht, jedoch ist der gleiche Widerstandsverlauf bezüglich der Lampenlage dargestellt. Die Angaben zur Statistik beziehen sich auf den gesamten Testdurchmesser von ca. 130 mm.

Dies ist auch anhand von Versuchen zur Reproduzierbarkeit der Dotierung zu ersehen. In der Abbildung 5.33 sind Dotierversuche zusammengestellt, die mit identischen Prozessrezepten durchgeführt wurden.

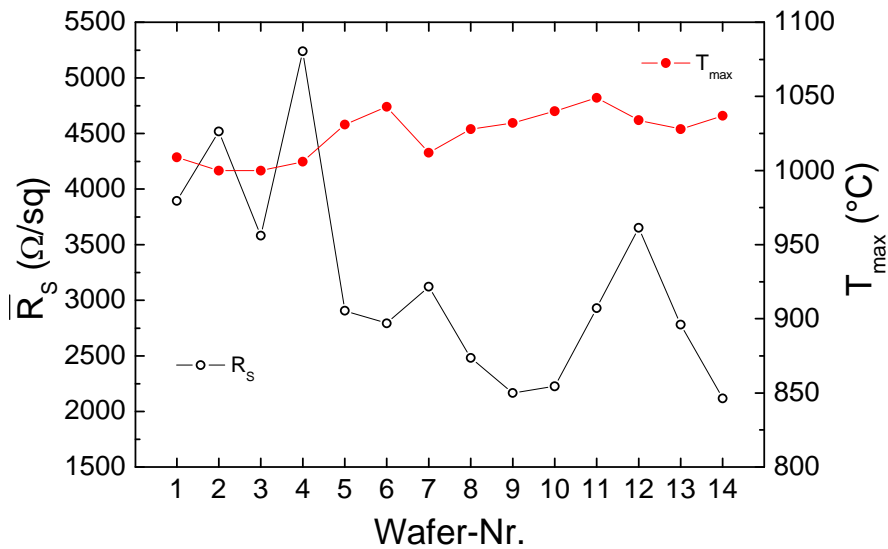


Abbildung 5.33: Reproduzierbarkeit der Dotiererergebnisse für 14 gleichartig durchgeführte Dotierprozesse. Dargestellt sind die Mittelwerte der gemessenen Flächenwiderstände in einem homogener dotierten Bereich von 75 mm Durchmesser in der Mitte der Wafer und die für diese Versuche aus dem aufgewendeten Lampenleistungen ermittelten Maximaltemperaturen. (Die Versuche wurden an verschiedenen Tagen durchgeführt, die Verbindungslinien symbolisieren keinen Zusammenhang, sondern sollen nur den Vergleich zu den Prozesstemperaturen erleichtern.)

Hieraus ist folgendes abzuleiten. Obwohl die identischen Rezepte verwendet wurden, sind die für die jeweiligen Versuche aufgewendeten Lampenleistungen unterschiedlich. Da diese für die Ermittlung der Prozesstemperatur maßgebend sind, ergeben sich die in der Abbildung dargestellten Temperaturvariationen. Die genaue Ursache hierfür ist nicht ganz klar. Ein Grund ist der oben beschriebene Einfluss von geringen Unterschieden in der Waferrandlage bezüglich der Temperaturerfassung. Da die Wafer jeweils automatisch transportiert wurden, waren hierbei jedoch kaum größere Abweichungen zu erwarten.

Die Waferreihenfolge in der Abbildung entspricht der zeitlichen Aufeinanderfolge der Versuche. Bei gleicher Solltemperatur ist ein Trend zu höheren Prozesstemperaturen festzustellen. Einen unbestimmten Einfluss kann eine Lampenalterung darstellen. Von den Lampen ist nicht bekannt, wie viele Betriebsstunden sie in Einsatz waren. Wenn für gleiche Strahlungsleistungen infolge Alterung eine höhere Leistungsaufnahme erforderlich würde, sollte dieser Trend wegen der gewählten Kalibriermethode keiner wirklichen Temperaturerhöhung entsprechen. Das gleiche wäre für eine fortschreitende Kontaminierung der Quarzglasplatte der Fall; sichtbare Kontaminationen wurden jedoch nicht festgestellt. Diese Effekte verdeutlichen die Beschränktheit einer Kontrolle der Prozesse durch eine einfache Steuerung von vorgegebenen Leistungswerten (open-loop) im Gegensatz zum geschlossenen Regelkreis (closed-loop-Verfahren).

Somit sind durch die Anlage selbst Grenzen für eine gute Reproduzierbarkeit von Prozessen gegeben. Aus der Abbildung wird jedoch ersichtlich, dass die Nichtreproduzierbarkeit der Prozesstemperaturen nicht allein für die Schwankungen der Prozessergebnisse verantwortlich sein kann. Es ergibt sich aus dieser Auswertung kein eindeutiger Zusammenhang zwischen der Temperatur und dem Mittelwert der Flächenwiderstände. Offenbar ist der gewählte Prozessverlauf nicht stabil genug gegenüber geringen (meist kaum kontrollierbaren) Variationen von Randbedingungen. Einen wichtigen Einfluss kann dabei auch die Reihenfolge der Wafer in einer unmittelbaren Abfolge von hintereinander ausgeführten Prozessen darstellen. Eingehendere Untersuchungen dazu erforderten einen hohen Wafereinsatz und wären für

diese Anlage und diesen Prozess jedoch kaum von großem Nutzen. So wurde dies nicht weiter verfolgt.

Prozessvariante 2

Für weitere Dotierexperimente wurde versucht, Prozesse zu entwickeln, die auch im Rahmen der beschränkten Anlagenmöglichkeiten günstigere Ergebnisse erbringen. Solche Prozesse sollten stabiler bezüglich geringer Variationen von Randbedingungen sein. Darüber hinaus wurde die Leistungskorrektur der Lampen für die folgenden Versuche verwendet, um die Homogenität der Dotierung zu verbessern. Damit ergaben sich außerdem, wie in Kapitel 5.2.3 beschrieben, deutlich höhere Prozesstemperaturen. Das Eintreiben in der statischen Sauerstoffatmosphäre zeigte bereits, dass geringe Sauerstoffdrücke von 2,5 mbar für ein erfolgreiches Dotieren genügen. Daraufhin wurde für die folgenden Versuche in diesem Schritt mit kontinuierlichem Sauerstofffluss gearbeitet, von dem, wie bei den Oxidationsversuchen gezeigt, homogenere Prozessergebnisse erwartet wurden. Ein maximaler stationärer Sauerstoffdruck von ca. 5 mbar wurde erreicht und verwendet.

Die mit der Prozessvariante 2 durchgeführten Experimente zeigten in der Tat wesentliche Verbesserungen in Homogenität, Reproduzierbarkeit und auch bezüglich der erzielbaren deutlich geringeren Flächenwiderstände. Einflüsse von Prozessvariationen konnten somit bewertet werden. Eine Auswahl der in den folgenden Diskussionen näher beschriebenen Experimente ist in der Tabelle 5.4 angegeben. Hierin sind außerdem die aus den Lampenleistungen ermittelten Maximaltemperaturen während des Eintreibens und die mittleren Flächenwiderstände und Standardabweichungen in Prozent vom Mittelwert angegeben. Diese wurden hier aus einem Testbereich von 120 mm Durchmesser in der Mitte des Wafers ermittelt.

Zur Verdeutlichung der Homogenität der Prozessergebnisse sind zusätzlich zu den Mittelwertangaben der jeweiligen Flächenwiderstände in Tabelle 5.4 Flächenwiderstandsverläufe für ausgewählte Waferdurchmesser in der Abbildung 5.34 dargestellt. Deren Lage entspricht denen der Oxidationsversuche.

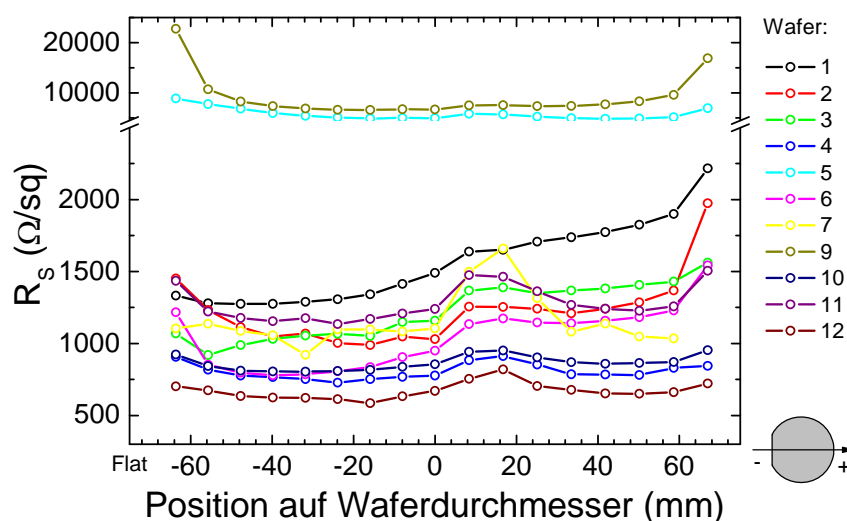


Abbildung 5.34: Darstellung des Verlaufs des Flächenwiderstands über dem Waferdurchmesser (rechts unten symbolisiert) für die links bezeichneten Wafer mit Wafernummern entsprechend Tabelle 5.4. Die Lage der Verläufe ist wie bei Oxidationsversuchen.

Tabelle 5.4: Ausgewählte Dotierexperimente der Untersuchungen mit Prozessvariante 2 des Zweistufenprozesses. Die Lampenheizung erfolgte mit Leistungskorrektur, vor dem Eintreiben erfolgte eine Vorheizphase von $T_{\text{Soll}} = 550\text{ °C}$ für 60 s, die Gasdrücke wurden durch stationären Gasfluss eingestellt, außer beim Eintreiben von Wafer 1. Die Ermittlung der maximalen Prozesstemperatur erfolgte aus der Auswertung der aufgewendeten Lampenleistungen, die Angaben zum Flächenwiderstand beziehen sich auf einen Testdurchmesser von 120 mm (47 Messpunkte). Die Erklärung der Abkürzungen von Prozessvariationen ist unten angegeben.

Wafer	Ausheizen	Belegen	Eintreiben				Flächenwiderstand	
	Medium	Medium	Medium	$t_{\text{Dot.}}$ (s)	T_{Soll} (°C)	T_{max} (°C)	\bar{R}_s (Ω/sq)	σ_{R_s} (%)
1	P	P	O-stat.	60	900	1153	1459	13,0
2	P	P	O	60	900	1158	1099	8,2
3	P	P	O	60	900	1155	1119	12,0
4	-	P	O	60	900	1175	769	5,1
5	H	P	O	60	900	1153	4321	9,5
6	P	-	O	60	900	1158	912	14,9
7	-	P	O	60	900	1163	1091	8,8
8	-	-	O	60	900	1163	-	-
9	P	P	-	-	-	-	8911	9,2
10	-	P	O	10	900	1165	835	4,5
11	-	P	O	10	850	1089	1219	6,7
12	-	P	O	10	950	1203	638	5,2
13	-	P	O	10	900	1157	847	6,5
14	-	P	O	10	900	1153	883	5,8
15	-	P	O	10	900	1173	774	6,5
16	-	P	O	10	900	1156	814	5,8
17	-	P	V	10	900	1173	-	-
18	-	P	O	10	900	1158	858	3,3
19	-	P	O	20	900	1153	757	5,7
20	-	P	O	40	900	1160	694	6,0

P	2,5 mbar 50 vpm PH_3 in H_2 , bei Ausheizen: $T_{\text{Soll}} = 900\text{ °C}$ für 60s, bei Belegung: $T_{\text{Soll}} = 550\text{ °C}$ für 60 s	O-stat. 40 mbar O_2 , statisch, kein Gasfluss O 5 mbar O_2
H	3 mbar H_2 bei Ausheizen: $T_{\text{Soll}} = 900\text{ °C}$ für 60s,	V kein Prozessgas, Druck $< 0,1$ mbar

Die in der Tabelle 5.4 und in Abbildung 5.34 aufbereiteten wichtigsten Ergebnisse enthalten wesentliche Informationen zu verschiedenen Einflüssen auf den Dotierprozess, die im Folgenden zusammengefasst werden sollen.

Zunächst soll der Vergleich der Ergebnisse von Wafer 4 mit den Wafern 2 und 5 erfolgen, wobei die Prozesse letzterer Ausheizschritte wie in Prozessvariante 1 enthielten. Beim Wafer 5, der in reiner Wasserstoffatmosphäre ausgeheizt wurde, ist ein deutlich höherer Flächenwiderstand zu verzeichnen, was für eine geringe oder wenig effiziente Dotierung spricht. Offenbar war hier die dem Ausheizen folgende Belegung sehr ineffizient oder auch diese in der folgenden sauerstoffhaltigen Eintreibatmosphäre zu flüchtig. Auch beim Ausheizen in der Wasserstoffatmosphäre mit dem geringen Phosphinanteil (Wafer 2 und wie generell bei Prozessvariante 1) ist sowohl die Dotierung geringer als auch inhomogener als bei Wafer 4, der ohne diesen Hochtemperatur-Ausheizschritt prozessiert wurde.

Das Ausheizen in Wasserstoff stellt hier somit keine positive Konditionierung der Oberfläche durch Entfernen von Oxidschichten dar, sondern bewirkt einen nachteiligen Effekt (insbeson-

dere für eine nachfolgende kontrollierte Belegung) durch Ätzen der Oberfläche in sauerstoffarmer Atmosphäre. Solches Oberflächenätzen durch die Umwandlung von Silicium zu gasförmigen SiO wurde für einen ähnlichen Kurzzeitprozess von Downey et al. beschrieben [Downey '98]. Es führt zu einer erhöhten Oberflächenrauigkeit, die für eine kontrollierte Adsorption von PH₃ abträglich ist. Beim Ausheizen in der phosphinhaltigen Wasserstoffatmosphäre ist dieses Ätzen mit einer beginnenden Dotierung aus der Gasphase verbunden, wie sie auch bei einer einstufigen Gasphasendotierung erfolgt. Mit einer anschließenden Belegung führt dies zwar zu einem geringeren Flächenwiderstand als beim Ausheizen in reiner Wasserstoffatmosphäre, jedoch ist dieser Prozess noch inhomogener und somit schwer zu kontrollieren, wie bei den Untersuchungen mit Prozessvariante 1 gezeigt wurde.

Die beginnende Gasphasendotierung konnte mit dem Wafer 9 gezeigt werden, bei dem nur das Ausheizen und die anschließende Belegung durchgeführt wurde, ohne dass ein Eintreibeschritt in der sauerstoffhaltigen Atmosphäre stattfand. Danach war ein sehr hoher Flächenwiderstand noch messbar¹, der für eine sehr geringe Dotiereffizienz spricht, wie in den Ausführungen zu dem einstufigen Dotierverfahren beschrieben.

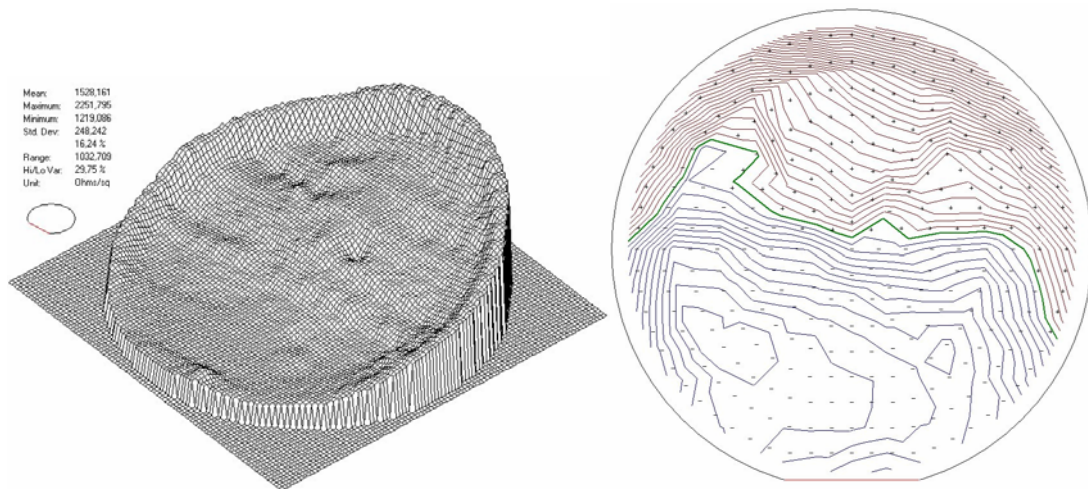
Im Vergleich zeigt der Wafer 4 hingegen, bei dem nur die Belegung bei geringerer Temperatur und danach das Eintreiben erfolgte (d.h. die Prozessvariante 2 wie eingangs vorgestellt durchgeführt wurde), einen deutlich geringeren Flächenwiderstand, der zudem noch gleichmäßiger über der Waferoberfläche war. Hier trat offenbar kaum ein Ätzen der Oberfläche auf, was positiv für eine kontrollierte Belegung ist. Dabei ist hinzuzufügen, dass diese Belegung nicht wie angestrebt bei 550 °C erfolgte, wofür die effektivste dissoziative PH₃-Adsorption auf Silicium erwartet wurde (siehe Kapitel 3.1.3). Sie erfolgte vielmehr aufgrund der Temperaturerfassungsfehler bei Temperaturen von ca. 800 bis 700 °C (siehe Kapitel 5.2.2 mit Tabelle 5.3). Noch günstigere Belegungen sollten somit für optimierte Belegungstemperaturen zu erwarten sein.

Das nachfolgende Eintreiben konnte nun in reinem Sauerstoff ohne weitere Dotierstoffzufuhr erfolgen, war somit nicht mehr abhängig von schwer kontrollierbaren Dotiergaswechselwirkungen mit der Oberfläche. Dass der Sauerstoff wiederum erforderlich war, zeigt das Ergebnis von Wafer 17, bei dem mit gleicher Belegung und versuchtem Eintreiben ohne die sauerstoffhaltige Atmosphäre keine Dotierung messbar war. Die Dotierung blieb auch aus, wenn kein Eintreiben erfolgte (Dies ist nicht in der Tabelle dargestellt.) Bei der geringen Temperatur des Belegungsschrittes ist auch keine nennenswerte Eindiffusion zu erwarten. Außerdem wurde mit Wafer 8 geprüft, ob eventuell allein durch den Eintreibeschritt ohne die Dotiergaszufuhr ein Dotiereffekt messbar ist, wie er etwa durch den Eintrag von verbliebenen Dotierstoffanteilen aus vorangegangenen Prozessen möglich wäre. Dies konnte jedoch nicht gefunden werden, so dass bei diesem Schritt von einem reinen Eintreiben ohne Dotierstoffzufuhr zur Oberfläche auszugehen ist.

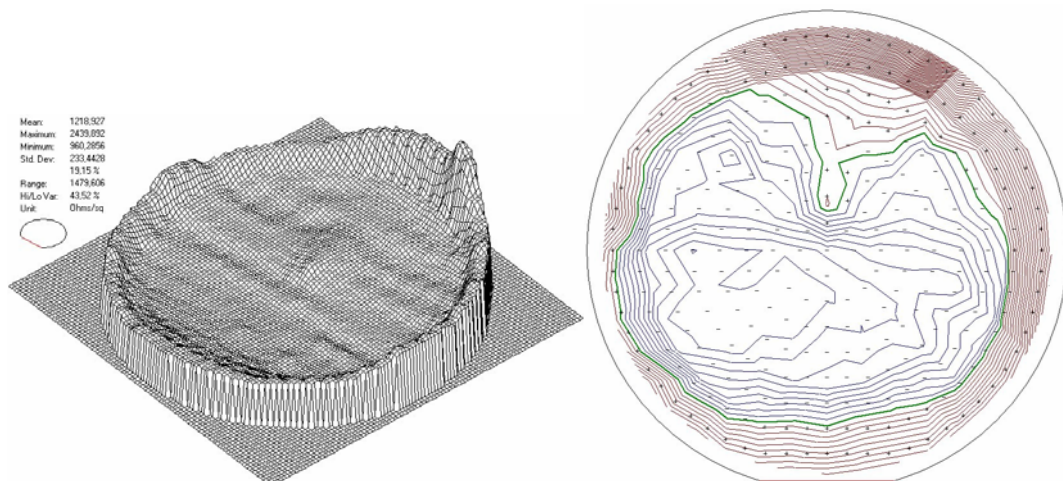
Mit einem Vergleich von Wafer 4 mit Wafer 7, bei denen jeweils die gleichen Prozessabläufe erfolgten, soll an dieser Stelle auf ein Problem der Prozessreproduzierbarkeit hingewiesen werden. Wafer 7 war der erste Wafer, der nach einer längeren Pause von 5 Tagen prozessiert wurde. Er ist geringer und deutlich inhomogener dotiert worden. Das Gleiche wurde auch für den Prozess mit Ausheizen beobachtet (Vergleich von Wafer 2 mit Wafer 3 als erstem Wafer nach Prozesspause). Wahrscheinlich genügte dafür eine geringe Kontamination über Kammerlecks, die in den vorliegenden Fällen zu einem Druckanstieg auf ca. 1,5 bzw. 2,5 mbar führte. Offenbar muss in solchen Fällen für das Erreichen reproduzierbarer Ergebnisse die Kammer durch Ausheizprozesse konditioniert werden.

¹ Bei den anfänglichen Versuchen mit konventionellen Dotierverfahren mit 50 vpm PH₃ in H₂ wurde sogar keinerlei messbare Dotierung gefunden, da mit geringeren Temperaturen prozessiert wurde.

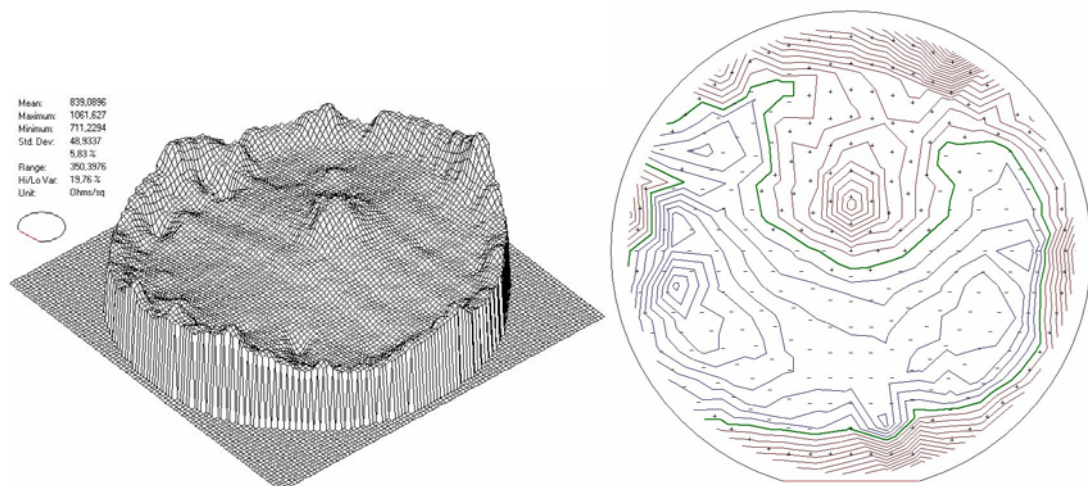
Eine weitere Diskussion von Variationen ausgewählter Prozessparameter in der Prozessvariante 2 (Wafer 10 bis 20) erfolgt weiter unten.



Wafer 1: mit Ausheizen in H_2/PH_3 und Eintreiben statisch bei 40 mbar O_2



Wafer 2: mit Ausheizen in H_2/PH_3 und Eintreiben stationär bei 5 mbar O_2



Wafer 10: ohne Ausheizen, nur Belegung in H_2/PH_3 und Eintreiben stationär bei 5 mbar O_2

Abbildung 5.35: Flächenwiderstände in Abhängigkeit vom Ort auf der Waferoberfläche in einer räumlichen Gitterdarstellung, die Höhe entspricht einem Maß für den jeweiligen Flächenwiderstand, und jeweils dazugehörige „Höhenliniendarstellung“ mit Linien gleichen Flächenwiderstands für drei verschiedene Wafer (siehe Tabelle 5.4).

Die Abhängigkeit der Dotierhomogenität von den gewählten Prozessen soll anhand der Abbildung 5.35 verdeutlicht werden. Jeweils dargestellt sind die Flächenwiderstände in Abhängigkeit von dem Ort auf der Waferoberfläche in einer räumlichen Gitterdarstellung, die eine R_S -Topographie darstellt, wobei die Höhe ein Maß für den jeweiligen Flächenwiderstand ist. Die entsprechende „Höhenliniendarstellung“ mit den Linien gleichen Flächenwiderstands ist jeweils links mit angegeben.

Mit Wafer 1 ist ein Vergleich eines Dotierergebnisses aus Prozessvariante 1 angegeben. Trotz der hier erfolgten Lampenkorrektur wurde dieser Wafer sehr inhomogen dotiert, der Flächenwiderstand fällt zum Flat hin deutlich ab. Bei den Oxidationsversuchen, die auch unter statischen Bedingungen reproduzierbar durchgeführt werden konnten, zeigte sich ein Abfall der Oxiddicke zum Flat hin, was auf kühlere Bereiche dort hinwies. Dies ließe dann höhere Flächenwiderstände erwarten. Wie bereits eingangs zur Prozessvariante 1 beschrieben, beeinflussen wohl kaum kontrollierbare zufällige Effekte, die auch von der Prozessierung als erster Wafer nach einer Prozesspause herrühren, die Homogenität und Reproduzierbarkeit äußerst nachteilig.

Bei Wafer 2 wird der wie bei den Oxidationsversuchen beschriebene Temperatureinfluss deutlich. Hier ist auch eine Welligkeit in der Flächenwiderstandsverteilung sichtbar, die durch die stabförmige Anordnung der Lampen bedingt ist. Am Waferrand ist der Flächenwiderstand deutlich höher, was auf einen kühleren Bereich durch erhöhte Abstrahlung hinweist. Hier erweist sich von Nachteil, dass keine Möglichkeit vorhanden ist, durch einen Guardring diesen Effekt zu minimieren. In der Abbildung von Wafer 2 ist rechts oben außerdem die Stelle der Waferauflage auf das Pin-Thermoelement als deutliche Erhöhung des Flächenwiderstands auf der Oberseite erkennbar. Dieses war bei allen dotierten Wafern mehr oder weniger deutlich ausgeprägt. Als Grund hierfür ist auch eine wie bereits bei den Oxidationsversuchen beschriebene lokale Temperaturenminderung zu vermuten. Etwas schwächer ausgeprägt konnte dies zumeist auch für die beiden anderen Auflagestellen auf die Quarzstifte gefunden werden. Weiterhin ist eine lokale Erhöhung des Flächenwiderstands nahe der Wafermitte zu finden. Diese Erhöhung war bei allen dotierten Wafern zu finden (sofern sie nicht bei sehr inhomogen dotierten Wafern, wie bei Wafer 1, überdeckt wurde). Die Ursache für diese Erhöhung ist nicht ganz klar, zumal bei den Wafern, die oxidiert wurden, an dieser Stelle eine lokale Erhöhung der Oxiddicke vorhanden war (siehe Abbildung 5.22/Abbildung 5.23), die als lokal heißere Stelle interpretiert wurde. Dies sollte also einen geringeren Flächenwiderstand zur Folge haben. Über dem Wafer ist an dieser Stelle an der Reaktoroberseite ein kleiner Flansch zur Druckmessung vorhanden, dieses „Loch“ in der Kammerwand stellt damit eine Störung im Absorptions- und Reflexionsvermögen gegenüber den benachbarten gekühlten Flächen dar. Eine lokale Überhitzung während des Adsorptionsprozesses könnte eine geringere Belegung zur Folge gehabt haben, was zu einem höheren Flächenwiderstand führte. Wie bei den Auswertungen der Dotierprofile gezeigt wird, sind an den Stellen des geringeren Flächenwiderstands sowohl weniger Dotierstoff nachweisbar, als auch flachere Profile vorhanden.

Die Darstellung des Wafers 10 ist nun ein Beispiel für ein Ergebnis aus Prozessvariante 2. Insgesamt ist die Verteilung deutlich homogener, was auch der Vergleich der statistischen Angaben für den gesamten Wafertestbereich zeigt. Die Erhöhung des Flächenwiderstands nahe der Wafermitte ist jedoch auch hier vorhanden.

Bei der Analyse der Flächenwiderstandsverteilungen wird deutlich, dass mit geeignet durchgeführten Dotierprozessen der beschriebenen Verfahren auch ein Temperaturmonitoring von Prozessanlagen durchführbar sein sollte. Eine Auswertung ist genauer und einfacher durch-

föhrbar als Schichtdickenmessungen am Ellipsometer. Auch ein Monitoring für kürzeste Prozesszeiten sollte damit möglich sein.

Zusammengefasst lassen sich aus der Durchführung der Prozessvariante 2 die folgenden wichtigsten Schlüsse ziehen.

- Eine erfolgreiche Belegung der Siliciumoberfläche mit Phosphor in einer Atmosphäre aus gering konzentriertem Phosphin in Wasserstoff ist ohne einen vorherigen Ausheizschritt bei Temperaturen oberhalb von 900 °C möglich. Die mittels der nasschemischen Verfahren gereinigten Siliciumoberflächen genügen damit den Anforderungen für eine erfolgreiche dissoziative Adsorption von Phosphin.
- Aus dieser Belegung ist im stationären Sauerstofffluss von 5 mbar eine homogenere und reproduzierbarere Dotierung als mit Prozessvariante 1 möglich. Der Sauerstoff ist aber auch hier erforderlich, um einen weitgehenden Verlust von Phosphor durch Desorption zu verhindern und um Oberflächenätzen zu vermeiden.

5.4.3 Flächenwiderstand in Abhängigkeit von Prozessparametern

Durch eine Variation ausgewählter wichtiger Prozessparameter der Prozessvariante 2 wurden deren Einflüsse auf den Flächenwiderstand ermittelt. Die Ergebnisse sind im Folgenden dargestellt (siehe auch Tabelle 5.4).

Temperatur des Eintreibeschrittes

Die Temperatur während des Eintreibens besitzt einen bedeutenden Einfluss auf den Flächenwiderstand der dotierten Proben. Die Ergebnisse der Experimente bei verschiedenen Eintreibetemperaturen sind in der Abbildung 5.36 dargestellt. Geringere Flächenwiderstände wurden mit höheren Dotiertemperaturen erzielt.

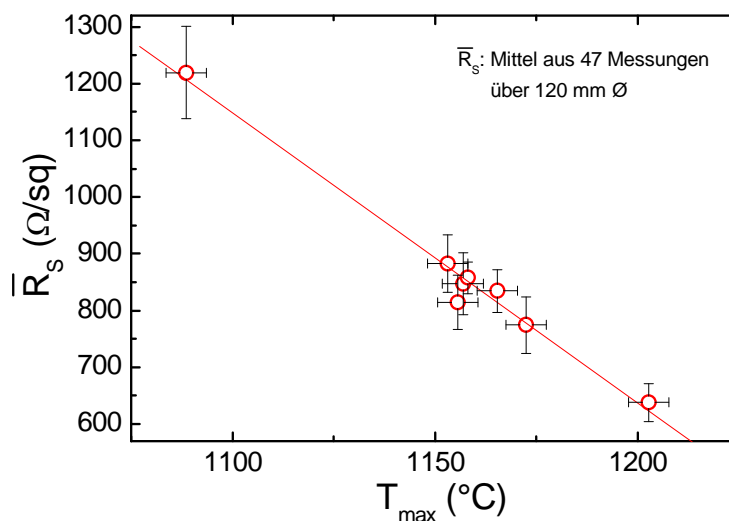


Abbildung 5.36: Flächenwiderstand (Mittelwert, siehe Tabelle 5.4) in Abhängigkeit von der Maximaltemperatur des Eintreibeschrittes für gleiche Eintreibzeit von 10 s. Die Gerade wurde aus linearer Regression ermittelt und ermöglicht die Abschätzung der Streuung der Prozessergebnisse.

In dieser Darstellung sind die korrigierten Maximaltemperaturen des Eintreibeschrittes dargestellt und die jeweilig erzielten mittleren Flächenwiderstände (120 mm Testdurchmesser), wobei als erste Näherung für deren Zusammenhang in dem dargestellten Temperaturbereich das Ergebnis der linearen Regression mit angegeben ist. Zunächst wurde eine scheinbar große Streuung der Ergebnisse für Prozesse, die alle bei gleicher Solltemperatur von 900 °C durchgeführt wurden, gefunden. Jedoch zeigt die Darstellung, dass nach Korrektur aus den tatsächlich aufgewendeten Lampenleistungen, die Messwertstreuung überwiegend infolge der tatsächlich während der Prozesse vorhandenen Temperaturunterschiede auftrat und somit durch die Probleme der Reproduzierbarkeit der Prozesstemperatur der Anlage begründet ist. Der Dotierprozess nach der Prozessvariante 2 ist hingegen bei besserer Kontrolle der Prozess-temperatur reproduzierbar durchzuführen, er ist insbesondere stabiler, als die oben beschriebene Variante 1.

Dauer des Eintreibeschrittes

Der Flächenwiderstand hing von der Haltezeit t_{Dot} des Hochtemperaturschrittes während des Eintreibens bei 5 mbar Sauerstoff ab, wie für das Dotieren bei gleicher Solltemperatur (900 °C) in Abbildung 5.37 gezeigt wird. Wie erwartet sinkt der Widerstand für längere Dotierzeiten. Die jeweilige tatsächliche Prozess-temperatur (T_{max}), die für diese Prozesse keine größeren Abweichungen voneinander aufwies und in der Darstellung mit angegeben ist, zeigt, dass diese Verringerung aufgrund der längeren Eintreibzeit erfolgt sein musste.

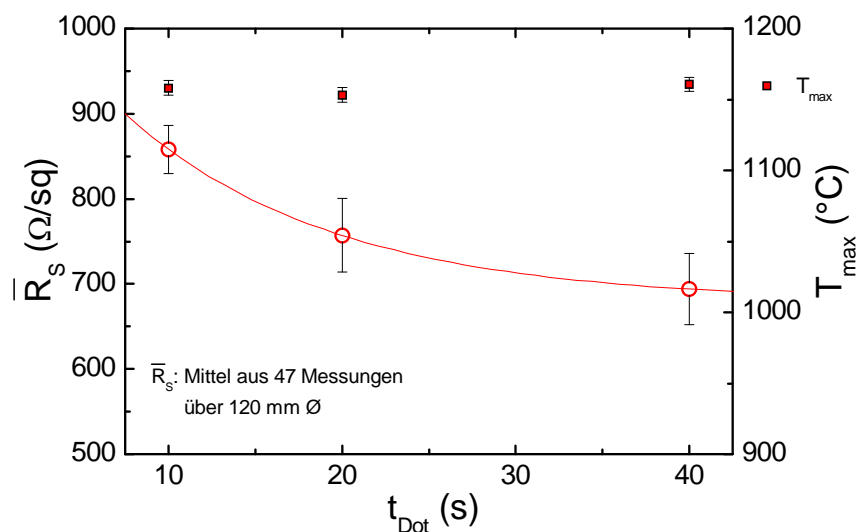


Abbildung 5.37. Flächenwiderstand in Abhängigkeit von der Eintreibzeit bei gleicher Temperatur (mit angegeben) in 5 mbar Sauerstoff. Die Angabe der Verbindungslinie erfolgte hier zur günstigeren Darstellung des Zusammenhangs, eine funktionelle Abhängigkeit kann aus den Ergebnissen noch nicht abgeleitet werden.

Allerdings zeigten die SIMS-Messungen (siehe unten), dass sich mit längerer Dotierzeit die pn-Übergangstiefen vergrößerten. Sehr kurze Eintreibzeiten sind somit für flache pn-Übergänge von herausragender Bedeutung.

Dauer der Belegung

Ein wichtiges Ergebnis ist, dass der Flächenwiderstand nicht kritisch von der Belegungszeit während des Adsorptionsschrittes abhängt. In der Tabelle 5.5 sind verschiedene Dotier-

versuche zusammengestellt, die unter gleichen Eintreibbedingungen, jedoch mit unterschiedlichen Belegungsbedingungen prozessiert wurden.

Tabelle 5.5: Übersicht über Dotierversuche mit verschiedenen Belegungsschritten und gleichen Eintreibbedingungen nach Prozessvariante 2, Angabe von Haltezeit während der Belegungstemperatur, maximaler Temperatur während des Eintreibens (ermittelt aus aufgewendeten Lampenleistungen) und mittlerem Flächenwiderstand (120 mm Testdurchmesser).

Wafer	Belegen	Eintreiben	Flächenwiderstand
Nr.	t_{Belegung} (s)	T_{max} (°C)	\bar{R}_s (Ω/sq)
A	60	1153	951,5
B	1	1158	848,4
C	10	1155	767,0
D	30	1175	766,8
E	60 ^{*1}	1153	783,6
F	10 ^{*2}	1158	766,8

^{*1} Wafer E: Variante 1 der Belegung: kein Dotiergasfluss während der Abkühlrampe von der Belegungstemperatur,

^{*2} Wafer F: Variante 2 der Belegung: nur 20 s Spülen im Dotiergasfluss (statt 60 s) und kein Dotiergasfluss während der Abkühlrampe von der Belegungstemperatur.

Die Abbildung 5.38 zeigt nun die erzielten mittleren Flächenwiderstände in Abhängigkeit von der Haltezeit bei der Belegungstemperatur¹. Um den Einfluss der Temperaturvariationen der Anlage abzuschätzen, sind die Ergebnisse außerdem dem in Abbildung 5.36 gefundenen Zusammenhang zwischen maximaler Eintreibtemperatur und mittlerem Flächenwiderstand in Abbildung 5.39 gegenübergestellt.

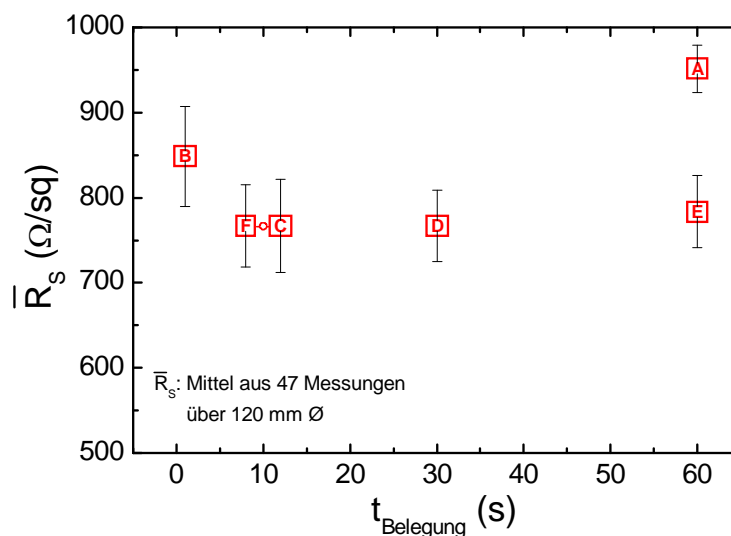


Abbildung 5.38. Flächenwiderstand in Abhängigkeit von der Zeit der Phosphineinwirkung bei gleicher Belegungstemperatur und gleichen Eintreibbedingungen. Für 10 s Belegungszeit sind zwei Datenpunkte mit demselben Widerstandswert eingezeichnet (Waferbezeichnungen entsprechend Tabelle 5.5).

¹ Zur Ermittlung der wahren Temperatur während der Belegung siehe Abbildung 5.11 und Tabelle 5.3.

Bei Wafer A handelt es sich wahrscheinlich um eine erhöhte Abweichung aufgrund seiner Prozessierung als erster Wafer in dieser Versuchsreihe. Sonst liegt die Streuung innerhalb der mit der Anlage möglichen. Das heißt, dass kein eindeutig auf eine unterschiedliche Belegung zurückzuführender Unterschied des Flächenwiderstands für eine Variation der Dauer der Belegung von 1 s bis zu 60 s festgestellt werden kann. Außerdem war es unerheblich, ob der Dotiergasfluss während des Abkühlens von der Belegungstemperatur aufrechterhalten wurde oder nicht oder ob ein vorheriges Spülen der Kammer mit Dotiergas von 20 oder 60 s erfolgte (siehe Wafer E und F). Dies bedeutet, dass innerhalb der betrachteten Zeitspanne offenbar bereits eine Sättigung der Belegung auftrat.

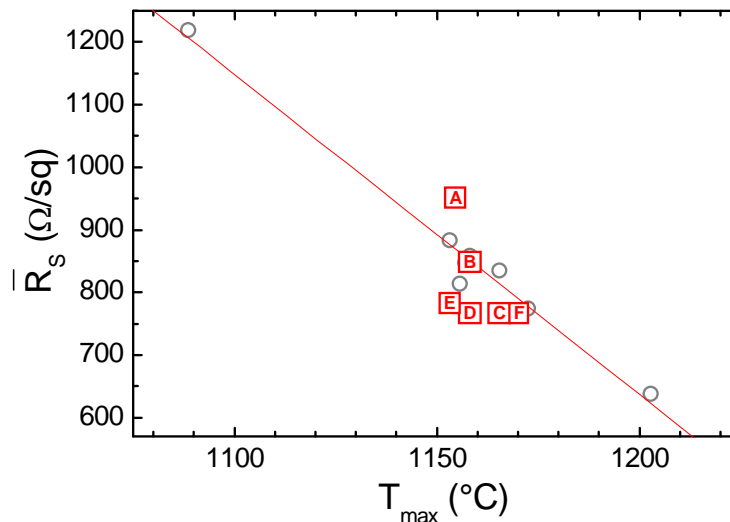


Abbildung 5.39: Einfluss der maximalen Eintreibetemperatur auf den Flächenwiderstand der unterschiedlich belegten Wafer (Waferbezeichnungen entsprechend Tabelle 5.5).

So zeigt sich, dass die Adsorption von Phosphor unter diesen Bedingungen selbstbeschränkt ist, wie in der Literatur beschrieben (z.B. in [Tsukidate '99]). Dort wird auch berichtet, dass die Sättigung der Oberflächenbelegung bereits mit sehr geringen PH_3 -Dosen von 12-16 L abgeschlossen ist, was den Einsatz des hoch verdünnten Dotiergases ermöglicht und damit günstig für die Vermeidung von Phosphorkontaminationen ist. Durch eine solche Selbstbeschränkung der Belegung ist diese darüber hinaus sehr unempfindlich gegenüber Variationen von Prozessbedingungen.

5.4.4 Flächenwiderstand und Oxidation

Die Messungen des Flächenwiderstands wurden im Reinraum an Luft durchgeführt. Dabei wurde beobachtet, dass die Messungen des Flächenwiderstands der dotierten Wafer scheinbar nicht reproduzierbar waren. Nach Wiederholungsmessungen zu späteren Zeitpunkten wurde ein Anstieg der gemessenen Flächenwiderstände bei sämtlichen dotierten Wafern festgestellt. In der Abbildung 5.40 ist dies für einen dotierten Wafer dargestellt. Die Messungen erfolgten in bestimmten Abständen nach der Entnahme aus der Vakuumschleuse nach dem Dotierprozess mittels der gewöhnlichen Messprozedur. Das heißt, der Wafer wurde für jede Messung auf den Vakuumchuck des Vierspitzenmessgeräts transportiert, ausgerichtet und die Messspitzen wurden für jede Messung neu aufgesetzt. Zwischen den Messungen wurde der Wafer an Luft belassen (für kurz aufeinander folgende Messungen) oder in der Waferbox aufbewahrt.

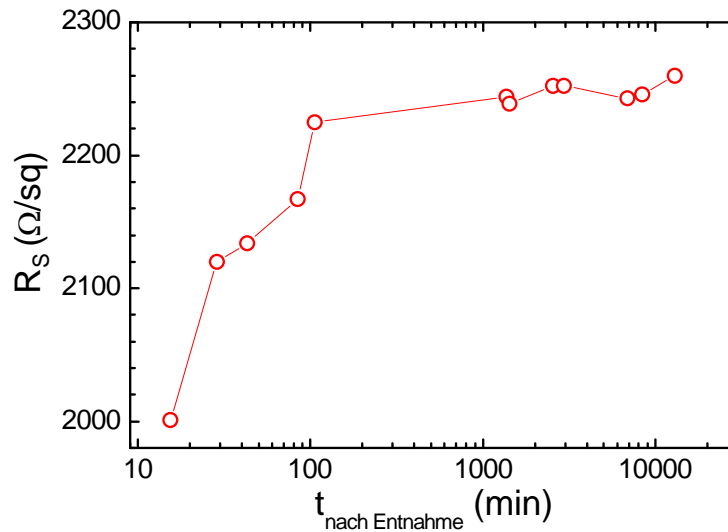


Abbildung 5.40: Flächenwiderstand gemessen in Wafermitte in Abhängigkeit von der Zeit des Waferkontakts mit Reinraumluft nach dem Dotierprozess (Wafer aus Prozessvariante 1).

In dieser Abbildung ist deutlich zu erkennen, dass während der ersten ca. 2 Stunden, die dieser Wafer nach dem Prozess der Luft ausgesetzt war, ein deutlicher Anstieg des Flächenwiderstands um mehr als 10 % des anfänglich gemessenen Wertes erfolgte, bevor ein weitgehend konstanter Wert erreicht wurde. Ein solcher Effekt an dotierten Wafern wurde bisher nicht beschrieben. Aufgrund der großen Bedeutung des Flächenwiderstandswertes für die Evaluierung und Quantifizierung von Dotierergebnissen und auch Prozessanforderungen (siehe einleitende Betrachtungen zur ITRS-Roadmap) sollten hierzu einige genauere Untersuchungen erfolgen.

Dafür wurde eine andere Messprozedur gewählt. Es ist bekannt, dass das Aufsetzen der Messspitzen auf die Waferoberfläche zu mechanischem Stress und sogar Beschädigungen der Oberfläche führt [Clarysse '02]. Bei zahlreichen wiederholten Messungen an derselben Stelle würde dies zu größeren Schäden und damit insbesondere bei den vorliegenden flachen Übergängen zu Messfehlern kommen, die den zu untersuchenden Effekt empfindlich stören könnten. Daher wurden für nun die Messspitzen nur einmal auf die Waferoberfläche aufgesetzt und verblieben dort für die Dauer der wiederholten Messungen. Ein weiterer Vorteil ergibt sich hierbei dadurch, dass keine Probleme der Reproduzierbarkeit durch geringe Veränderungen der Messposition der aufeinander folgenden Einzelmessungen auftreten können. Die Flächenwiderstandswerte wurden in festgelegten Zeiten ermittelt und gespeichert. Auch hier wurde die duale Messkonfiguration benutzt, und die Messungen erfolgten an Luft im Reinraum; der Wafer war dabei durch die Abdeckung des Messgeräts vor direktem Lichteinfall geschützt.

In der Abbildung 5.41 ist das Ergebnis einer solchen Messung dargestellt. Hierin ist deutlich der Anstieg des Flächenwiderstands in einer Art gestuften Kurve zu erkennen. Nach einem ersten Verlangsamten des Anstiegs folgt ein Bereich eines schnelleren Anstiegs, bevor dieser nach ca. 2,5 Stunden wieder abnimmt und sich der Flächenwiderstand einem konstanten Wert annähert.

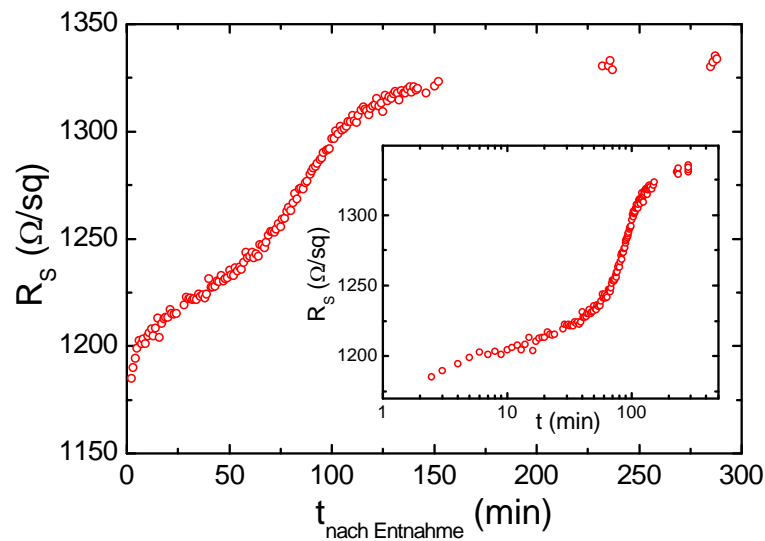


Abbildung 5.41 Anstieg des Flächenwiderstands von Wafer 3 nach der Entnahme aus der Vakuumschleuse nach dem Dotierprozess bei einer Messung mit aufgesetzten Messspitzen an Luft. In dem eingefügten Diagramm ist die Zeitskala zum Vergleich logarithmisch dargestellt.

Dieser Verlauf kann nicht zufällig sein, da die Einzelwerte keine große Streuung aufweisen. Es kann auch davon ausgegangen werden, dass dieser Anstieg nicht einfach aufgrund von sich verschlechternden Kontakteigenschaften der Messspitzen oder von Veränderungen des Ausbreitungswiderstands durch Oxidationen unter den Spitzen aufgetreten ist. Solche Widerstandseffekte wurden durch die Kelvin-Ausführung des Messsystems eliminiert, die eine praktisch stromlose Spannungsmessung ermöglicht [Schroder '98]. Außerdem wurde ein Anstieg, wie zuvor beschrieben, auch mit den gewöhnlichen Messungen, für die die Messspitzen jeweils erneut aufgesetzt wurden, beobachtet¹.

Es wurde vermutet, dass der Anstieg des Flächenwiderstands mit einem Anstieg der Oxiddicke, wie er bei in Kapitel 5.1.3 für undotierte Wafer nach der Eingangsreinigung beschrieben wurde, in Zusammenhang steht. Daher wurde im Folgenden versucht, dieses Wachstum auch an dotierten Wafern, bei denen mit einem HF-Dip die Oxidschicht entfernt wurde, nachzuweisen und nach einer Wiederholung der HF-Behandlung auch einen Flächenwiderstandsanstieg zu finden.

Für die Wafer, die mittels der Prozessvariante 2 dotiert wurden, ließ sich die Oxiddicke mit dem Ellipsometer wie bei den undotierten unter der Annahme des einfachen Schichtmodells nach Abbildung 4.9 bestimmen (siehe auch Kapitel 5.4.7). Die nun mit den beiden unabhängigen Verfahren kurz nacheinander auf demselben Wafer (Wafer 16) festgestellten Anstiege bei Kontakt mit der Reinraumluft sind in der Abbildung 5.42 gegenübergestellt.

¹ Im Allgemeinen ist es üblich, bei Vorhandensein einer dünnen Oxidschicht zu messen. Dabei durchdringen die Spitzen das Oxid. Dies ist wahrscheinlich ein Grund dafür, dass solch ein Anstieg noch nicht beschrieben wurde.

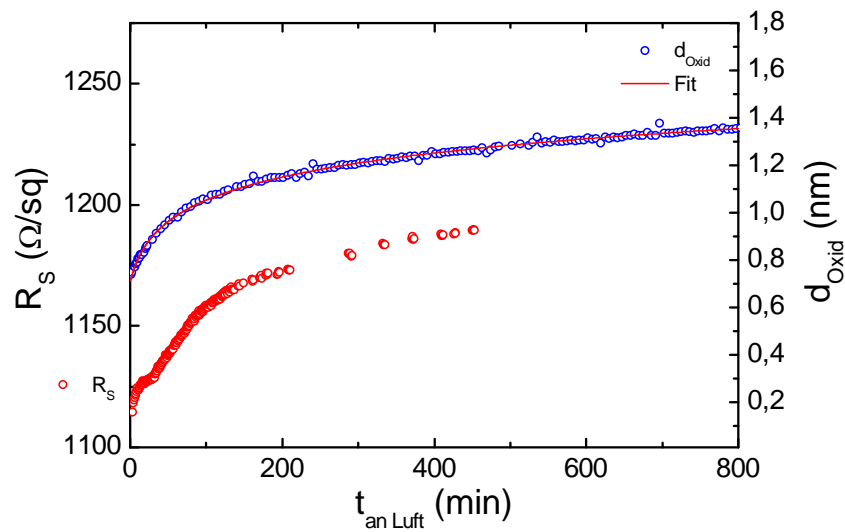


Abbildung 5.42: Vergleich des Anstiegs des Flächenwiderstands eines mit Prozessvariante 2 dotierten Wafers mit dem Anstieg einer mittels Ellipsometer gemessenen Schichtdicke desselben Wafers jeweils in Luft nach einer Oxidentfernung durch HF-Behandlung.

Hierin ist zu erkennen, dass der Anstieg der Oxiddicke mit der Zeit in ähnlicher Weise erfolgte, wie der Anstieg des Flächenwiderstands, der hiermit auch für einen dotierten Wafer nach einer HF-Behandlung nachgewiesen wurde. Im Anstieg der Oxiddicke ist jedoch keine ausgeprägte Stufe zu erkennen, wie sie beim Flächenwiderstandsanstieg zwar geringer als bei dem Wafer direkt nach dem Prozess jedoch noch deutlich ausgeprägt ist.

Es kann gefolgert werden, dass der Anstieg des Flächenwiderstands nicht aufgrund von speziellen Oberflächenbedingungen nach der Dotierung im zweistufigen Prozess aufgetreten ist. Der hier vorgestellte Wafer wurde mehrfach mit HF behandelt, und die ellipsometrischen Messungen zeigten, dass sich die Oberflächenstruktur wie die der undotierten Wafer für die Analyse der Spektren beschreiben lässt (siehe Kapitel 5.4.7). Außerdem zeigen die SIMS-Ergebnisse der HF-behandelten Wafer keine besonderen Oberflächenanlagerungen von Phosphor im Vergleich zu den dotierten Wafern, die ohne vorherige HF-Behandlung analysiert wurden. So muss der R_S -Anstieg aufgrund der geringen Tiefe der Dotierung aufgetreten sein und sollte auch für Proben mit entsprechender Dotierung auftreten, die auf andere Art als mit dem vorgestellten Dotierprozess geschaffen wurden. Damit sollten zeitabhängige Flächenwiderstandsmessungen ein relativ einfaches Untersuchen von Schichtwachstumsprozessen auf Siliciumoberflächen und von deren Auswirkungen auf die elektrischen Eigenschaften der Halbleiterstruktur gestatten.

Der Grund einer Beeinflussung der elektrischen Eigenschaften durch ein Schichtwachstum infolge einer Oxidation kann eine Deaktivierung von Dopanten und eine Verminderung der Übergangstiefe durch die Siliciumumsetzung der Oxidationsreaktion sein. Es können aber auch Veränderungen der Ausdehnung der Raumladungszone aufgrund von Akkumulation und Austausch von Oberflächenladungen infolge von Oberflächenanlagerungen auftreten. Für beide Effekte ist von Bedeutung, dass bei geringer Tiefe der Übergänge nahe der Oberfläche geringe Änderungen an der Oberfläche relativ große Auswirkungen auf den Flächenwiderstand besitzen, der damit ein empfindlicher Indikator für diese Prozesse ist.

Eine Möglichkeit der Auswertung der Beobachtungen und ein Aufzeigen von Zusammenhängen soll im Folgenden gezeigt werden. Es soll angenommen werden, dass das Schichtwachstum aufgrund einer Oxidation des Siliciums erfolgte. An dieser Stelle muss hinzugefügt werden, dass keine Vorkehrungen getroffen wurden, eine Adsorption von Kontaminationen

zu verhindern. Der Wafer war auf dem Messtisch des Ellipsometers der Reinraumluft ausgesetzt, so wie auch bei der Vierspitzenmessung. Da wie in Kapitel 4.4.1.1 beschrieben, eine eindeutige Ermittlung der Art des Schichtmaterials allein aufgrund der ellipsometrischen Parameter für die hier gewachsenen dünnen Schichten nicht möglich ist, kann auch anderes Schichtwachstum als durch eine Oxidation erfolgt sein, wie z.B. das von Kontaminationsschichten aufgrund von möglichen Verunreinigungen der Luft mit organischen Verbindungen [Olsen '89; Kasi '91; Saga '97]. Im Folgenden soll trotzdem einfach von Oxidwachstum gesprochen werden, da eine natürliche Oxidation von Silicium an Luft stattfindet und es in den hier vorgestellten Untersuchungen vorrangig um die Methode gehen soll, einen zeitabhängigen Oberflächenprozess, der mit einem nichtelektrischen Messverfahren beobachtet werden kann, dem Flächenwiderstandsanstieg gegenüberzustellen.

Das Schichtwachstum kann durch ein direkt-logarithmisches Wachstumsgesetz beschrieben werden. Solches Wachstum wurde z.B. von Fehlner und Mott für die natürliche Oxidation von Metallen und Halbleitern beschrieben, die netzwerkbildende Oxide aufweisen, wie es auch bei Silicium der Fall ist [Fehlner '70]. In der Abbildung 5.42 ist ein Fit nach diesem Gesetz entsprechend der Gleichung

$$d_{\text{Oxid}} = d_0 + A \ln(1 + B t_{\text{Luft}}) \quad (71)$$

angegeben, wobei $A = 0.152 \text{ nm}$, $B = 0.0861 \text{ min}^{-1}$, and $d_0 = 0.709 \text{ nm}$ die Fitparameter darstellen. Der Parameter d_0 wurde hierbei als Erweiterung zum originalen Wachstumsgesetz eingefügt, um eine anfängliche Schichtdicke (z.B. aufgrund einer Oxidation im DI-Spülwasser) zu beschreiben und auch um eine Unbestimmtheit bei der ellipsometrischen Messung der absoluten Dicke auszugleichen. Es kann festgestellt werden, dass dieser Fit den Anstieg gut beschreibt, es wurde ein Regressionsparameter von $R^2 = 0.99733$ bestimmt.

Es soll nun davon ausgegangen werden, dass das gleiche Wachstum auch während der Zeit des Flächenwiderstandsanstiegs erfolgte. Da die ellipsometrischen Messergebnisse des Schichtwachstums durch das logarithmische Gesetz gut beschrieben werden konnten, können damit die R_S -Werte des jeweiligen Messzeitpunktes mit der entsprechenden Schichtdicke infolge des Oxidwachstums in direkten Zusammenhang gebracht werden. Durch Umrechnung über die Messzeit ergibt sich eine Abhängigkeit des Flächenwiderstands von der Schichtdicke, wie sie in der Abbildung 5.43 dargestellt ist.

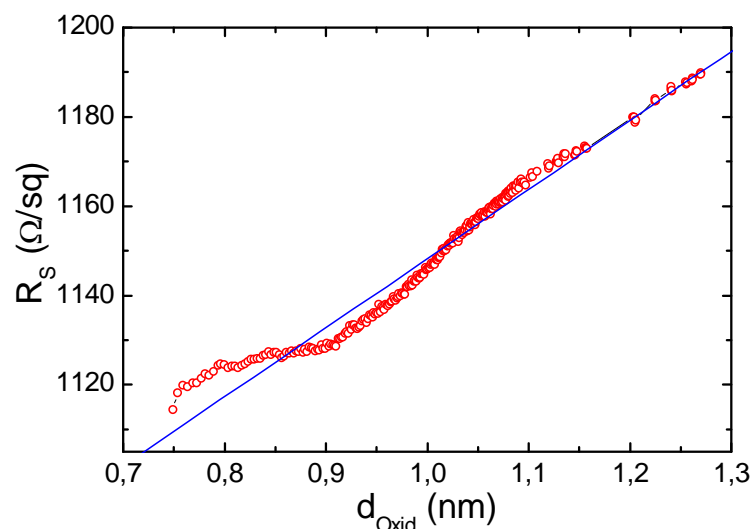


Abbildung 5.43: Flächenwiderstand in Abhängigkeit von der Dicke des wachsenden Oxids.

In dieser Abbildung ist zu erkennen, dass der Flächenwiderstand abgesehen von auffälligen Schwankungen linear mit dem Schichtwachstum ansteigt. Wenn die Differenz des tatsächlich gemessenen Flächenwiderstands von dem Wert gemäß dem linearen Fit¹ über der jeweiligen erwarteten Oxiddicke dargestellt wird, ergibt sich dar in der Abbildung 5.44 angegebene Verlauf.

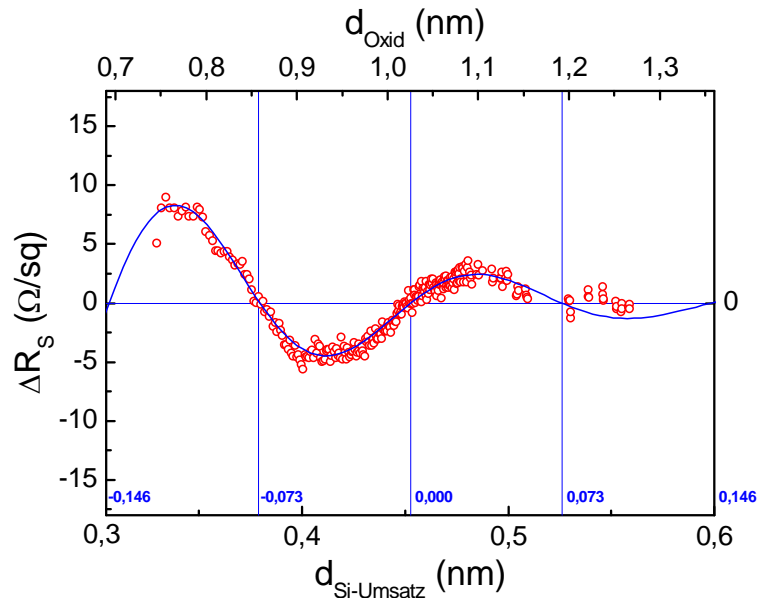


Abbildung 5.44: Differenz zwischen dem gemessenen Flächenwiderstand und dem linearen Fit aus Abbildung 5.43 über der erwarteten Oxidschichtdicke (oben) bzw. der Tiefe des Siliciumumsatzes durch die Oxidation. Die Gitterlinien kennzeichnen die Periodizität der Oszillationen, die dazu angegebenen Werte beziehen sich dabei auf die Tiefe des Siliciumumsatzes. Der eingezeichnete Verlauf einer gedämpften Sinusfunktion soll hier nur als Referenz dienen.

In dieser Abbildung sind die Abweichungen von der linearen Abhängigkeit deutlich sichtbar. Hierin ist außerdem angegeben, welche Tiefe des ursprünglichen Siliciums bei einer angenommenen Oxidation entsprechend der gewachsenen Oxiddicke umgesetzt wurde. Dabei wurde von einer gewöhnlichen Oxidation ausgegangen, bei der das Verhältnis von Tiefe des umgesetzten Siliciums zur Oxiddicke 0,44 beträgt. Auffällig ist nun, dass die Abweichungen Schwankungen entsprechen, deren Periodizität aufgrund des logarithmischen Wachstumsgesetzes der Schicht auf einer logarithmischen Zeitskala auftritt. Solche Oszillationen wurden für ein durch den Tunneleffekt bestimmtes Schicht-für-Schicht-Wachstum von natürlichen Oxiden auf Silicium beschrieben [Vlad '99]. Interessant ist, dass die in obiger Abbildung gefundene Periode der Schwankungen mit ca. 0,15 nm ungefähr dem Abstand der (100)-Ebenen in Silicium entspricht.

So könnten die Ursache der Schwankungen des Flächenwiderstands eigentlich Schwankungen des Wachstums der Schicht selbst sein, die mit dem einfachen logarithmischen Wachstumsgesetz nicht beschrieben werden konnten. So ist in der Tat bei entsprechender Bildung der Differenz von beobachtetem Wachstum und dem erwarteten nach dem einfachen logarithmischen Modell nach (71) der in der folgenden Abbildung 5.45 dargestellte Zusammenhang zu finden.

¹ Die Differenz bezeichnet also die Abweichung des gemessenen Flächenwiderstands von dem Wert eines Flächenwiderstands, der sich einstellen würde, wenn ein exakter linearer Zusammenhang zwischen Flächenwiderstand und Oxidwachstum gemäß dem idealen Wachstumsgesetz nach (71) bestünde.

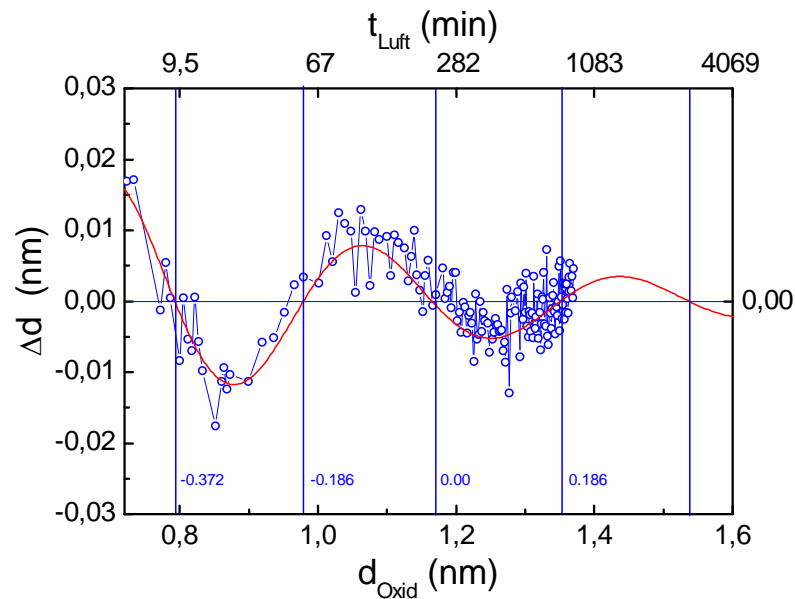


Abbildung 5.45: Differenz von der gemessenen Oxidschichtdicke und der nach dem logarithmischen Gesetz (71) zu erwartenden Schichtdicke (Oxidwachstum und Fit aus Abbildung 5.42) in Abhängigkeit von der erwarteten Schichtdicke. Einige offensichtlich stark gestreuten Messwerte wurden hier für eine bessere Anschaulichkeit nicht mit dargestellt. In der oberen Skala ist zum Vergleich die Oxidationszeit an Luft angegeben. Die Gitterlinien zeigen die Periodizität der Schwankungen auf der Oxiddickenskala. Der Verlauf der gedämpften Sinusfunktion ist nur als Referenz angegeben.

Obwohl sehr gering, sind auch hier Abweichungen zu dem einfachen Wachstumsgesetz in Form von regelmäßigen Schwankungen zu erkennen. Deren Periodizität entspricht dabei in etwa der in Abbildung 5.44 gefundenen Periode für die Dicke des Oxids beim Flächenwiderstandsanstieg. Dies kann ein Hinweis auf ein Schicht-für-Schicht-Wachstum an der Oberfläche sein. Obgleich diese Schwankungen sehr gering und mit dem Ellipsometer offenbar gerade noch messbar sind, hätten diese einen großen Einfluss auf den Flächenwiderstand. Dies würde bedeuten, dass die Wachstumsschwankungen gewissermaßen auf den Flächenwiderstandsanstieg abgebildet wurden. Eine einfache Messung des zeitlichen Verlaufes dieses Widerstands ließe somit Rückschlüsse auf Prozesse an der Siliciumoberfläche zu, die im Hinblick auf die sehr geringen Wachstumsraten mit anderen Methoden nur sehr schwierig zu detektieren sind.

Nicht ganz geklärt ist, wie ein solch geringes Wachstum von Oxid den Flächenwiderstand so erheblich beeinflussen kann. Bei Vergleich mit dem Dotierprofil (siehe Kapitel 5.4.5.3) und einer Abschätzung einer Flächenwiderstandserhöhung infolge einer Deaktivierung von Phosphor in dem durch die Oxidation umgesetzten Silicium fällt auf, dass ein solcher Einfluss viel geringer sein sollte. Offenbar spielen Oberflächenladungen im Zusammenhang mit einem Schichtwachstum an der Oberfläche die entscheidendere Rolle, so dass auch ein Wachstum von Adsorbatschichten und damit erfolgende Umladungen der Oberfläche der Grund für den beobachteten Flächenwiderstandsanstieg sein können. Bei weiteren HF-Behandlungen wurden jeweils gleichförmige Verschiebungen der R_S -Verläufe zu höheren Widerständen hin beobachtet. Auch dafür wurde nach Auswertung des Dotierprofils gefunden, dass allein der Materialabtrag durch die Oberflächenätzungen nicht der Grund für den beobachteten Flächenwiderstandsanstieg sein konnte.

Eine weitere Aufklärung dieses Effekts soll Gegenstand weiterer Untersuchungen sein. Wichtig an dieser Stelle sind die Folgen für die Quantifizierung der Dotierergebnisse und die Vergleichbarkeit der Prozesse. Die Flächenwiderstandswerte der Untersuchungen der Prozessergebnisse beziehen sich auf den Bereich, in dem keine wesentliche Änderung mehr

beobachtet wurde, das heißt, es wurde mindestens drei Stunden nach Entnahme aus der Vakuumschleuse oder nach einer HF-Behandlung gewartet, bevor die Flächenwiderstandsmessungen durchgeführt wurden.

5.4.5 Dotierstoffverteilung im Silicium

5.4.5.1 Vergleich SIMS-Messungen mit verschiedener Primärionenenergie

Um Informationen über die Verteilung des eindiffundierten Phosphors im Silicium zu erhalten, wurden SIMS-Analysen durchgeführt. Wie im Kapitel 4.4.2.2 bereits ausführlich dargelegt, stellt dabei jedoch die Analyse von sehr flachen Profilen eine besondere Herausforderung dar. In der Abbildung 5.46 sind zum Vergleich Messergebnisse einer Standard-SIMS-Messung bei 14,4 keV Cs^+ -Primärionenenergie denen einer Messung mit geringer Primärionenenergie von 3,5 keV gegenübergestellt.

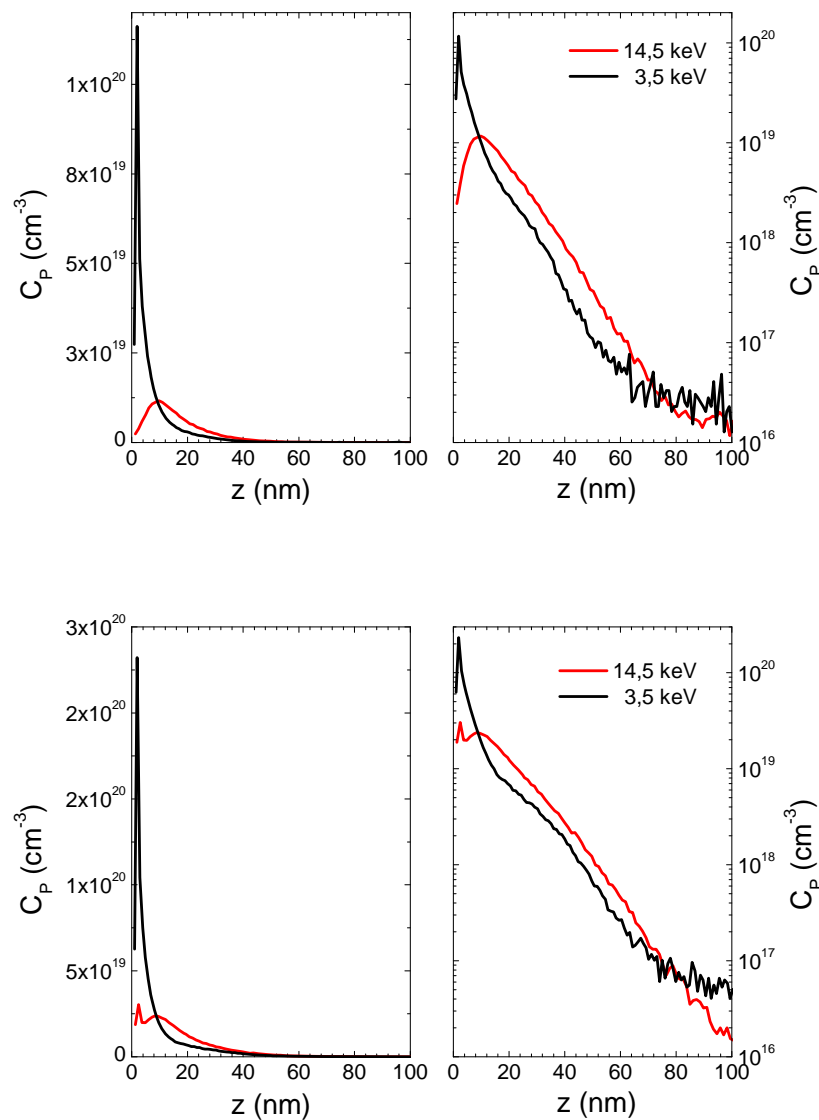


Abbildung 5.46: Vergleich von mit SIMS gemessenen Profilen von Phosphor in Silicium für zwei verschiedene Primärionenenergien und zwei verschiedene Dotierungen, links ist die ermittelte Phosphorkonzentration jeweils linear aufgetragen.

Hieran ist deutlich zu erkennen, dass mit der Standardmessung bei diesen Profilen bereits keine zuverlässigen quantitativen und qualitativen Analysen mehr möglich sind. Sowohl die Tiefe der Profile, als auch deren Form und besonders deren Anfangsbereich knapp unterhalb der Oberfläche sind offenbar deutlich verändert gegenüber den Profilen, die aus Messungen bei geringerer Primärionenenergie gewonnen wurden. Diese Messungen wurden jeweils an denselben Proben durchgeführt, die Veränderungen in den Profilen sind somit durch die Messungen selbst entstanden. Offenbar ist es durch das Cs^+ -Sputtern mit hoher Energie zu einem starken Mischeffekt gekommen, bei dem Phosphor aus einem Bereich hoher Phosphorkonzentration nahe der Oberfläche in tiefere Regionen eingemischt und dann dort detektiert wurde. Die Informationen zu einem Phosphorpeak an der Oberfläche, wie mit der geringen Primärionenenergie gemessen, gehen dabei im Prinzip vollständig verloren, nur für etwas höhere Peakkonzentrationen (in der Abbildung unten) wurde auch ein kleiner Peak bei den Standardmessungen mit höherer Energie erhalten.

Bei Vergleichsmessungen von Proben mittels IMS 4F-E6 ($E_{\text{Cs}} = 3,5 \text{ keV}$) und der ATOMIKA 6500 ($E_{\text{Cs}} = 4,5 \text{ keV}$) wurden fast identische Profilverläufe des Phosphors im Silicium gefunden, hier wird deutlich, dass (zumindest für den Teil des Profils unterhalb des Oberflächenpeaks) die Profilform nicht mehr von der Primärionenenergie abhängt und somit den tatsächlichen Verlauf repräsentiert. (Eine Diskussion des Oberflächenpeaks folgt im Kapitel 5.4.5.4.)

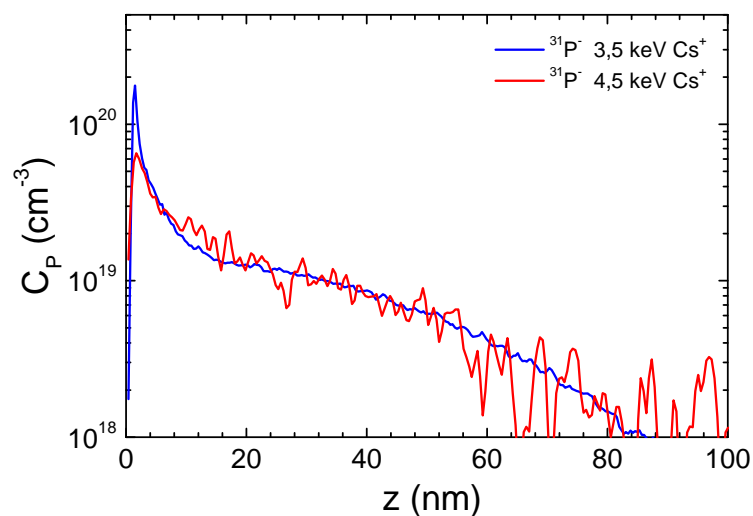


Abbildung 5.47: Vergleich von SIMS-Profilen einer Probe, die an verschiedenen Anlagen unter Verwendung von verschiedenen Primärionenenergien aufgenommen wurden.

Aus der Abbildung 5.47 ist auch zu ersehen, dass die Messungen mit dem ATOMICA 6500-System größere Messwertschwankungen für den betrachteten Konzentrationsbereich aufwiesen. Hier wird ein Problem der spektrometrischen Detektion von Phosphor in Silicium mit dem Quadrupol-Massenspektrometer deutlich. Das Messsignal wird hier durch einen hohen Untergrund aus $^{30}\text{Si}^1\text{H}^-$ überlagert, was mit diesem System aufgrund einer geringeren Massenauflösung von $^{31}\text{P}^-$ nicht zu unterscheiden ist, so dass das Detektionslimit hierbei bereits bei einer P-Konzentration von ca. $2 \cdot 10^{18} \text{ cm}^{-3}$ ist.

Zusätzlich zur Änderung der Profilform ist durch die Messungen bei der hohen Primärionenenergie auch eine Verminderung der detektierten Gesamtdosis des Phosphors im Silicium festzustellen. Eine solche Dosis oder Flächenkonzentration kann aus der Integration der

Profile gewonnen werden. Für alle jeweils mit beiden Energien ermittelten Profile ist diese Verminderung aufgetreten, wie in der Abbildung 5.48 dargestellt ist.

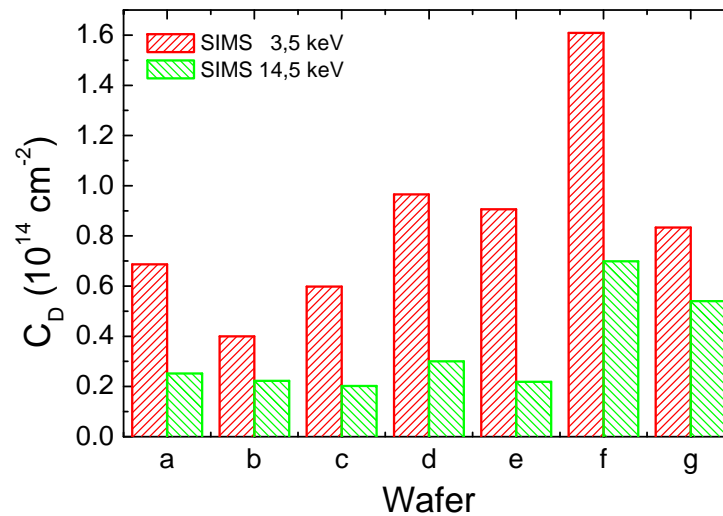


Abbildung 5.48: Vergleich der aus den SIMS-Profilen ermittelten Dosis in Abhängigkeit von der für die Messung verwendeten Primärionenenergie.

Durch den Beschuss mit Cs^+ -Ionen der hohen Energie ist es damit nicht nur zu einer Einmischung des Oberflächenphosphors gekommen, sondern wahrscheinlich auch zu unkontrolliertem Verlust von Phosphor, der nicht mit detektiert werden konnte.

Somit zeigt sich, dass die SIMS-Messungen unter Verwendung von 14,4 keV Cs^+ -Primärionenenergie für eine Analyse der Dotierergebnisse aus den betrachteten Prozessen nicht geeignet ist und für die folgenden Auswertungen nur Messungen mit einer Primärionenenergie kleiner als 4,5 keV verwendet wurden. Der Nachteil von geringeren Energien ist jedoch ein höherer Messaufwand durch eine geringere Sputter- und Detektionsrate, was auch zu höherem Rauschen im Messsignal und damit zu einem höherem Detektionslimit führt [Oberbeck '03].

5.4.5.2 Vergleich der Ergebnisse der unterschiedlichen Prozessvarianten

Im Folgenden sollen kurz Ergebnisse der Dotierung mit Prozessvariante 1 vorgestellt und mit denen aus der verbesserten Variante 2 verglichen werden. Aufgrund der mangelhaften Prozessreproduzierbarkeit und auch wegen der ermittelten Flächenwiderstandsinhomogenitäten der einzelnen Wafer sollen die Ergebnisse aus der Variante 1 nur qualitativ bewertet werden, da sinnvolle Schlussfolgerungen aus quantitativen Resultaten nur bei Ausschluss von zufälligen und unbestimmten Einflüssen auf das Prozessergebnis möglich sind.

In der Abbildung 5.49 ist ein Vergleich von Dotierprofilen, die von unterschiedlichen Stellen eines Wafers aus Prozessvariante 1 stammen, dargestellt. Die vor den SIMS-Analysen ermittelten Flächenwiderstände der jeweiligen Bereiche sind mit angegeben.

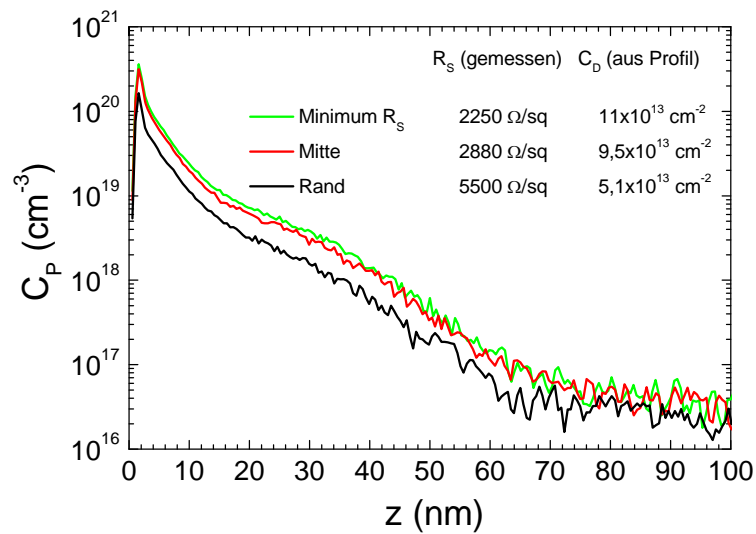


Abbildung 5.49: Vergleich der Dotierprofile (Konzentration von P in Si) von unterschiedlichen Bereichen eines Wafers, dotiert mittels Prozessvariante 1. Angegeben sind die Flächenwiderstände und die Flächenkonzentrationen von Phosphor von den entsprechenden Stellen.

Am Rand war dieser um mehr als den Faktor 2 höher, als im Bereich des geringsten Flächenwiderstands nahe der Mitte¹. Die Form der Profile aus den unterschiedlichen Bereichen ist ähnlich, jedoch unterscheiden sich die gemessenen Gesamtkonzentrationen deutlich voneinander. Die aus der Integration der SIMS-Profile ermittelten Flächendichten von Phosphor sind in Abbildung 5.49 mit angegeben. Der höhere Flächenwiderstand am Waferrand ist offenbar nicht nur eine Folge einer geringeren Eindringtiefe, sondern auch Folge einer geringeren Gesamtmenge von dort vorhandenem Phosphor, so dass auch die Belegung an dieser Stelle geringer gewesen sein sollte.

Bei einem Vergleich der Dotierergebnisse von Prozessvariante 1 und 2 (siehe Abbildung 5.50) fällt auf, dass auch hier jeweils qualitativ gleiche Profilverläufe erhalten wurden.

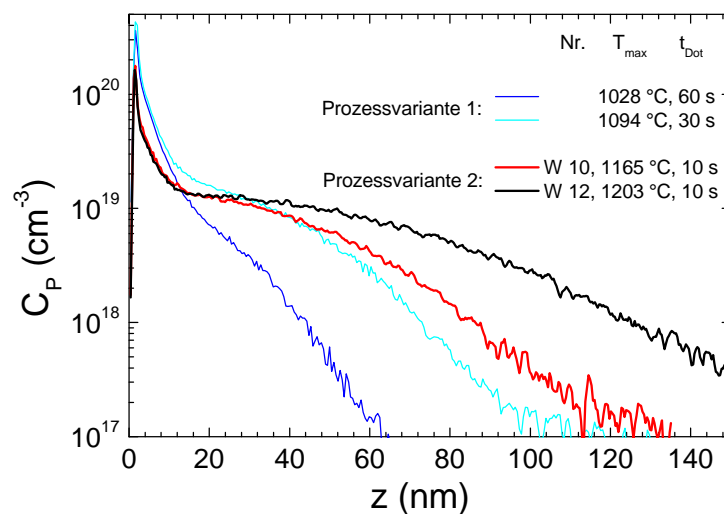


Abbildung 5.50: Vergleich von Phosphorprofilen in Silicium von mit den zwei unterschiedlichen Prozessvarianten dotierten Wafern.

¹ Der Verlauf des Flächenwiderstands ist ähnlich dem, wie in der Abbildung 5.31 gezeigt.

Die Ergebnisse von Prozessvariante 1 weisen einen höheren Anteil von Phosphor im Bereich des Oberflächenpeaks auf. Für die Prozessvariante 2, die bei höheren Eintreibetemperaturen durchgeführt wurde, ergeben sich tiefere Profile, auffällig ist jedoch, dass relativ unabhängig von der Eintreibetemperatur und damit der Tiefe, die maximale Konzentration des eindiffundierten Phosphors im Profilbereich unterhalb des Oberflächenpeaks ungefähr bei $1,3 \cdot 10^{19} \text{ cm}^{-3}$ liegt.

Dies legt den Schluss nahe, dass zumindest für die gesamte Dauer der hier betrachteten Eintreibbedingungen noch ein Depot von Phosphor an der Oberfläche vorhanden war, aus dem mit fortschreitender Eindiffusion in das Silicium dieser daraus nachgeliefert werden konnte.

In der Abbildung 5.51 sind SIMS-Ergebnisse von vier Dotierprozessen gegenübergestellt. Die Wafernummern entsprechen denen aus Tabelle 5.4.

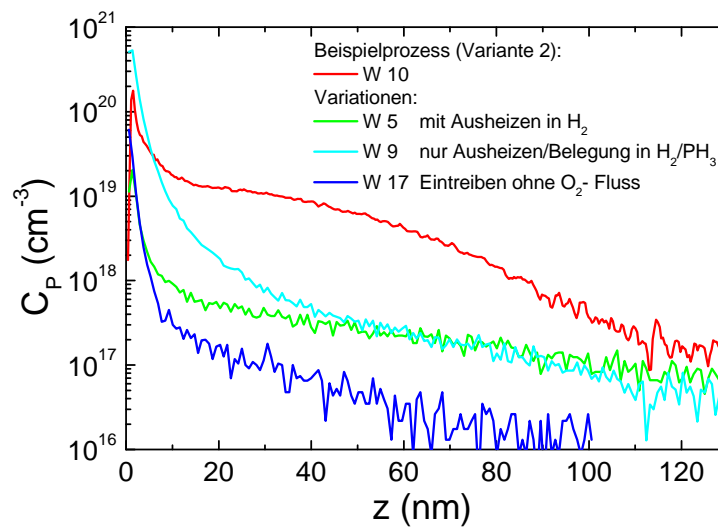


Abbildung 5.51: Einfluss von Variationen von Prozessbedingungen der Dotiervariante 2 auf die Verteilung des Phosphors im Silicium (Cs^+ -Primärionenergien: W10 - 3,5 keV, W5, W9, W 17 - 0,5 keV).

Im Vergleich zu dem verbesserten Prozess nach Variante 2, mit dem geringere Flächenwiderstände erzielt werden konnten (Wafer 10, wie zuvor diskutiert), ist hieraus folgendes abzuleiten.

Nach gleicher Prozessführung wie bei Wafer 10, jedoch Eintreiben ohne Sauerstofffluss (Wafer 17), ist kaum Phosphor in das Silicium eindiffundiert und auch nur ein geringerer Oberflächenpeak nachzuweisen. Bei der Eintreibetemperatur von über $1100 \text{ }^\circ\text{C}$ ohne die Sauerstoffzugabe ist offenbar als Folge eines Oberflächenätzens fast die gesamte Belegung desorbiert, ohne dass nennenswert Phosphor in das Silicium eindringen konnte. Entsprechend war auch kein Flächenwiderstand bei diesem Wafer messbar.

Ein Ausheizen in Wasserstoffatmosphäre vor der Belegung (Wafer 5) führte zu einer geringeren Belegung der Oberfläche. Nach dem Eintreiben unter gleichen Bedingungen wie bei Wafer 10 sind hier sowohl ein geringerer Oberflächenpeak als auch eine geringere Menge eindiffundierten Phosphors nachzuweisen.

In Wafer 9 wurde eine erheblich höhere Phosphorkonzentration als in Wafer 5 gefunden, obwohl er einen höheren Flächenwiderstand aufwies. Mit Wafer 9 wurde nur ein Ausheizen und anschließendes Belegen in Dotiergasatmosphäre durchgeführt. Das bedeutet, dass zwar

bei dem Hochtemperaturschritt bereits ein geringer Anteil des Phosphors in das Silicium diffundiert ist, der hohe gemessene Phosphoranteil im Bereich des ausgeprägten Oberflächenpeaks jedoch aus der nachfolgenden Belegung stammen muss. Dieser Anteil konnte damit noch nicht aktiv gewesen sein. Jedoch konnte dieser in einem nachfolgenden Eintreibeschritt in O₂-Atmosphäre aktiviert werden, wie ein Vergleich der Flächenwiderstände von Wafer 9 mit Wafer 2 und 3, bei denen die gleiche Vorbehandlung erfolgte, zeigt. Für eine weitere Diskussion der Oberflächenbelegung sei auf die nachfolgenden Kapitel verwiesen.

5.4.5.3 Quantitative Auswertung der Phosphor-Tiefenprofile im Silicium

Für quantitative Auswertungen aus SIMS-Profilen und deren Vergleich mit den Ergebnissen aus den elektrischen Messungen wurden nur die Wafer gewählt, die mittels der technologisch günstigeren Prozessvariante 2 dotiert wurden.

Die Ergebnisse der SIMS-Untersuchungen der drei bei unterschiedlicher Eintreibetemperatur dotierten Wafer (siehe Tabelle 5.4) sind in Abbildung 5.52 dargestellt. Die Profile wurden unter Verwendung der Cs⁺-Primärionenenergie von 3,5 keV aufgenommen.

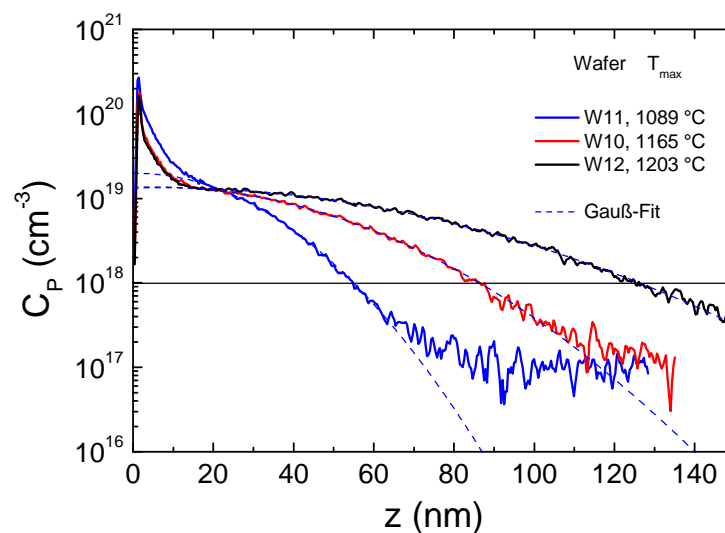


Abbildung 5.52: Mittels SIMS gemessene Konzentrationen von Phosphor in Silicium nach der Zwei-Stufen-Dotierung bei verschiedenen Eintreibetemperaturen. Die entsprechenden angenäherten Gauß-Verläufe sind gestrichelt dargestellt.

Der Verlauf der Phosphorkonzentration im Silicium konnte für den Bereich ab ca. 20 nm unter der Oberfläche durch eine Gauß-Funktion

$$C_p = C_{p_0} + 2 \frac{A}{w\sqrt{2\pi}} e^{-\frac{(z-z_0)^2}{w^2}} \quad (72)$$

angenähert werden. Die Fitparameter sind hierbei A , w und z_0 . Der Wert für die Konstante C_{p_0} wurde auf $7,5 \cdot 10^{13} \text{ cm}^{-3}$ festgelegt. Er entspricht der im Siliciummaterial aufgrund von Verunreinigungen vorhandenen Phosphorkonzentration und wurde aus den Angaben für übliche Phosphorgrenzwerte in undotiertem Silicium ermittelt. Eine Veränderung dieses Wertes um mehrere Größenordnungen hat jedoch praktisch keine relevanten Auswirkungen auf den Fit, bzw. die übrigen Fitparameter, solange er nur mindestens eine Größenordnung unterhalb des Detektionslimits der SIMS-Profile bleibt. Die aus einer üblichen numerischen Fitprozedur

erhaltenen Gauß-Funktionen sind in der Abbildung als gestrichelte Linien dargestellt. Man erkennt die sehr gute Übereinstimmung mit den gemessenen Profilen im betrachteten Bereich. Dieser Gauß-Verlauf deutet darauf hin, dass die Bedingungen des Eintreibens aus einer erschöpflichen Quelle während der Diffusion vorgeherrscht hatten (drive-in).

Die Übergangstiefe z_j (definiert bei $1 \cdot 10^{18} \text{ cm}^{-3}$ und markiert in der Abbildung 5.52) wurde aus den SIMS-Profilen ermittelt¹. Die Flächendichte C_D von eingebrachtem Phosphor (entspricht der Dosis bei der Implantation) wurde aus der Integration der Gauß-Verläufe bestimmt. Für einen Vergleich der aus den gemessenen Phosphorkonzentrationen im Silicium zu erwartenden Flächenwiderstände mit den tatsächlich gemessenen wurden mittels der Simulationssoftware ATHENA und ATLAS (Fa. SILVACO) diese Widerstände für die gegebenen Konzentrationsverläufe für jeweils zwei Fälle berechnet:

1. Der gesamte Phosphor der SIMS-Profile ist aktiv.
2. Nur die Phosphorkonzentration, die durch die Gauß-Funktion beschrieben wird, ist aktiv.

Die Ergebnisse sind in Tabelle 5.6 zusammengefasst.

Tabelle 5.6: Zusammenfassung ausgewählter Ergebnisse von Dotierexperimenten mit dem Zwei-Stufen-Prozess. Die Ergebnisse wurden aus den gemessenen Phosphor-Konzentrationsverläufen (SIMS) und aus den dem jeweiligen Gauß-Fit entsprechenden Konzentrationen (Gauß) ermittelt. Einige Wafer wurden vor der SIMS-Analyse mit verdünnter Flußsäure (HF, 1%, 3 min) behandelt.

Wafer	Ergebnisse aus der Profilanalyse						Flächenwiderstand R_s			d_{Oxid} (nm)	
	T_{max} (°C)	$C_D(\text{SIMS})$ (10^{13} cm^{-2})	$C_D(\text{Gauß})$ (10^{13} cm^{-2})	$C_{OF}(\text{SIMS})$ (10^{19} cm^{-3})	$C_{OF}(\text{Gauß})$ (10^{19} cm^{-3})	z_j (nm)	D_P ($10^{13} \text{ cm}^2/\text{s}$)	gemessen (Ω/sq)	berech. SIMS (Ω/sq)		berech. Gauß (Ω/sq)
keine HF-Behandlung vor SIMS-Analyse:											
11	1089	11,2	5,6	26,9	1,98	55	2,4	1150	789	1082	2,11
10	1165	9,8	6,7	17,7	1,37	86	6,3	800	703	826	2,29
12	1203	12,3	9,6	16,3	1,34	126	12,5	620	518	570	2,48
Oxidentfernung vor SIMS-Analyse (einmal HF-Dip):											
20	1160	7,8	7,5	2,2	1,23	106	10,9	710	713	720	1,55
D	1175	7,6	7,1	6,6	1,25	91	6,2	821	778	789	1,66
Oxidentfernung vor SIMS-Analyse (6 x HF-Dip):											
15	1173	6,6	6,3	4,9	1,25	88	7,1	1080	862	866	1,77

Aus diesen Ergebnissen lassen sich nun die folgenden wichtigen Schlussfolgerungen ziehen.

Flächenwiderstand und Aktivierung

Der Phosphoranteil nahe der Oberfläche, der in den SIMS-Analysen als deutlicher anfänglicher Peak bei den nicht mit HF behandelten Wafern zum Ausdruck kommt, kann nicht zur aktiven Dotierung beigetragen haben. Die berechneten Flächenwiderstände ausgehend von den mit SIMS gemessenen Phosphorkonzentrationen, das heißt unter der Annahme, der gesamte mit SIMS gemessene Phosphor wäre aktiv, waren signifikant geringer als die,

¹ Da die ermittelten Gauß-Verläufe eine sehr gute Näherung für das Profil des eindiffundierten Phosphors aufweisen, wurden die Übergangstiefen unter Vorgabe von $C_P = 1 \cdot 10^{18} \text{ cm}^{-3}$ nach Umstellen der Fitfunktion (72) nach $z = z_j$ (für C_P) errechnet.

mittels Vierspitzenmessung tatsächlich gemessen wurden. Andererseits waren die Flächenwiderstände, die berechnet wurden, wenn ausschließlich eine Gauß-Verteilung des Phosphors angenommen wurde, nahe an den gemessenen Werten. Die Abweichungen, die hier noch auftreten, entstehen wahrscheinlich aufgrund der Schwierigkeit, die Lage der Grenzfläche aus den SIMS-Profilen genau genug zu bestimmen¹. Der Vergleich der Flächenwiderstände ist in der Abbildung 5.53 graphisch dargestellt. Zum Vergleich ist hier auch der in Abbildung 5.36 dargestellte Zusammenhang zwischen dem mittleren Flächenwiderstand und der maximalen Temperatur des Eintreibeschrittes mit angegeben. Die Abweichungen zu diesem Verlauf sind überwiegend damit zu begründen, dass die exakte Temperatur an der Stelle, von der die Proben für die Untersuchungen entnommen wurden, nicht bekannt war.

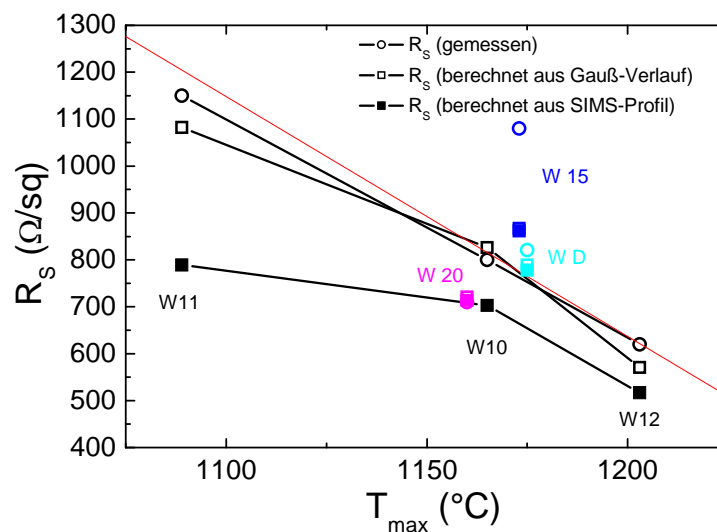


Abbildung 5.53: Abhängigkeit des gemessenen Flächenwiderstands von der maximalen Eintreibetemperatur im Vergleich zu den aus den SIMS-Profilen ermittelten Werten. (Schwarz dargestellt $t_{dot} = 10$ s, rot $t_{dot} = 40$ s und HF-Dip vor SIMS-Messung, die dünne Linie zeigt den Zusammenhang aus den Mittelwerten des Flächenwiderstands für 120 mm Testdurchmesser aus Abbildung 5.36).

Aus Abbildung 5.53 wird nun deutlich, dass der aus den mittels SIMS ermittelten Konzentrationen unter der Annahme der vollständigen Aktivierung des Phosphors zu erwartende Flächenwiderstand viel geringer ist, als der gemessene. Mit höherer Temperatur und damit wachsendem Anteil von eindiffundiertem Phosphor im Verhältnis zum gesamten detektierten nimmt der Unterschied ab. Die gemessenen und die aus den Gauß-Verläufen ermittelten Flächenwiderstände zeigen hingegen eine weitgehende Übereinstimmung, so dass geschlossen werden kann, dass der so eindiffundierte Phosphor auch tatsächlich aktiv ist.

Bestätigt wird dies auch anhand der Waferproben, bei denen eine Oxidentfernung vor den SIMS-Analysen durchgeführt wurde. In Abbildung 5.53 mit dargestellt sind auch die Ergebnisse der Wafer 15, 20 und D. Nach einmaliger HF-Behandlung, d.h. nach Entfernen einer hauptsächlich prozessbedingten Oxidschicht bei Wafer 20 und Wafer D liegen die drei Werte der Flächenwiderstände jeweils dicht beieinander. Dies bedeutet, dass der Oberflächenphosphor, durch die HF-Behandlung offenbar weitgehend entfernt werden konnte, so dass der Oberflächenpeak in den SIMS-Profilen nicht mehr so ausgeprägt war und die Gauß-Verläufe nun im Prinzip das gesamte Konzentrationsprofil repräsentieren. Ein Vergleich der SIMS-Profile für verschiedene HF-Behandlungen ist weiter unten in Abbildung 5.58 gezeigt, wo eine weitere Diskussion des markanten Oberflächenpeaks erfolgt. Die niedrigeren Flächen-

¹ Die Schwierigkeiten der *genauen Messung* der Flächenwiderstände der dotierten Wafer wurden bereits in Kapitel 5.4.4 beschrieben. Die hier verwendeten Ergebnisse stammen aus den Messungen, die durchgeführt wurden, nachdem der hohe Anstieg des Flächenwiderstandswerts vorüber war.

widerstände von Wafer 20 ergeben sich aufgrund der längeren Eintreibzeit von 40 s bei gleicher Temperatur. Wie unten gezeigt wird, kommt es hierdurch zu einer größeren Übergangstiefe aber auch zu einer gestiegenen Gesamtmenge eindiffundierten Phosphors.

Nach mehrmaligen Behandlungen mit HF, nach denen jeweils ein natürliches Oxid wuchs (Wafer 15), wurde auch nur ein sehr geringer Oberflächenpeak in dem SIMS-Profil gefunden, jedoch wurde nach den HF-Behandlungen jeweils ein höherer Flächenwiderstand gemessen (siehe Kapitel 5.4.4). Nach der sechsten Behandlung, nach der das SIMS-Profil ermittelt wurde, war der vorher gemessene Flächenwiderstand erheblich höher, als dies nach Auswertung der Phosphorkonzentration im Silicium zu erwarten wäre. Die genaue Ursache hierfür ist nicht geklärt.

Phosphorkonzentrationen

In der folgenden Abbildung 5.54 sind die ermittelten Konzentrationen in Abhängigkeit von der maximalen Eintreibtemperatur dargestellt. Hierin ist die Flächenkonzentration des aktiven Phosphors (rote Punkte und rote Linie) der Flächenkonzentration gegenübergestellt, die aus der gesamten gemessenen Phosphorkonzentration bestimmt wurde ($C_{D(SIMS)}$, schwarze Kreise). Weiterhin ist die Oberflächenkonzentration des aktiven Phosphors mit angegeben, die aus dem Maximum der Gauß-Funktion an der Oberfläche ermittelt wurde. Die Ergebnisse aus den Analysen der HF-behandelten Wafer sind durch farbige Umrandungen der Punkte hervorgehoben. Für diese wurden die Gesamtkonzentrationen nicht mit dargestellt, da durch die HF-Behandlung der Oberflächenphosphor entfernt wurde.

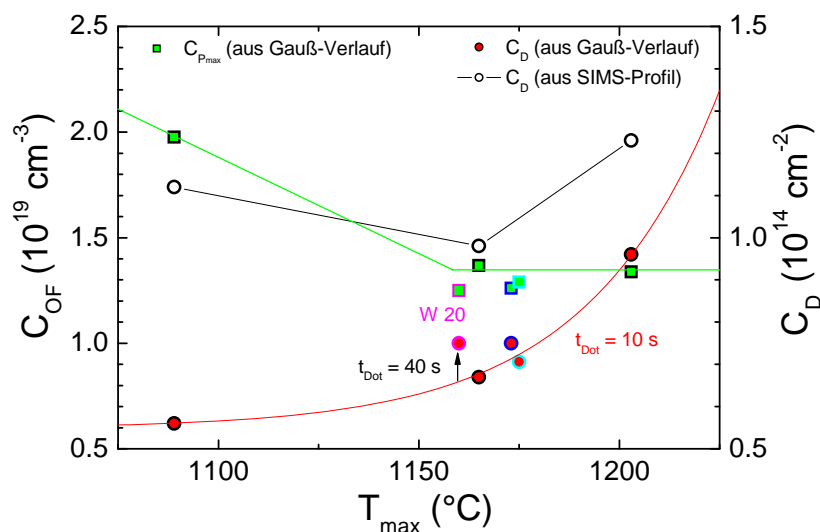


Abbildung 5.54: Zusammenfassende Darstellung der aus den SIMS-Profilen ermittelten Konzentrationen (Oberflächenkonzentration C_{OF} des aktiven Phosphors aus Gauß-Verlauf und Flächenkonzentration C_D [Dosis] aus Integration der gemessenen Profile und der Gauß-Verläufe). Die farbig umrandeten Punkte sind Ergebnisse von SIMS-Analysen von HF behandelten Proben (d.h. ohne Pile-up, siehe unten), die Farben der eingezeichneten Linien der ermittelten Zusammenhänge beziehen sich auf die Füllfarben der entsprechenden Datenpunkte.

Die Analyse dieser Ergebnisse ergibt nun, dass sich die Flächenkonzentration des aktiven Phosphors (rot: $C_{D(Gauß)}$), die aus der Integration der gaußförmigen Dotierprofile ermittelt wurde, mit steigender Dotiertemperatur erhöhte. Dies bedeutet, dass mit höherer Eintreibtemperatur mehr Phosphor auf aktive, substitutionelle Gitterplätze diffundiert ist. Für die Gesamtkonzentration des Phosphors (schwarz: $C_{D(SIMS)}$) konnte ein eindeutiger Zusammenhang nicht gefunden werden, aufgrund der Schwierigkeit einer genauen quantitativen Messung im Bereich des Oberflächenpeaks, worin ein Großteil des Phosphors noch vorhan-

den und nicht aktiv gewesen sein musste. Aus den Messungen ergibt sich jedoch ein Hinweis darauf, dass diese Konzentration nicht abhängig von der Eintreibetemperatur war und somit auch kein erhöhter Verlust von Phosphor bei höherer Eintreibetemperatur durch Desorption vorgelegen haben kann.

Für die Oberflächenkonzentration des aktiven Phosphors (grün) wurde gefunden, dass diese mit steigender Temperatur zunächst etwas geringer zu werden scheint, jedoch zumindest ab etwa 1150 °C nicht weiter fällt. Auch für den Wafer 20, der mit 4fach längerer Eintreibedauer prozessiert wurde, ist kein deutlicher Abfall dieser Oberflächenkonzentration festzustellen, obwohl die Konzentration des aktiven Phosphors deutlich höher war. Offenbar war die Oberflächenquelle von Phosphor für die hier untersuchten Eintreibbedingungen noch nicht „erschöpft“ und konnte noch Phosphor an aktive Gitterpositionen liefern. Eine genaue Beschreibung dieses Depots und der Vorgänge, die zu dieser Art Sättigung der Oberflächenkonzentration führten, ist aus den vorliegenden Messungen nicht zu erschließen. Die Festkörperlöslichkeit von Phosphor in Silicium sollte jedoch noch nicht der limitierende Faktor gewesen sein, da sie für den untersuchten Temperaturbereich mindestens eine Größenordnung höher ist (siehe Abbildung 3.1).

Zusammengefasst ergibt sich auch hier, dass zumindest für die untersuchten Eintreibbedingungen eine Quelle von Phosphor an der Oberfläche noch vorlag, dort jedoch nicht aktiv war und für eine effektivere Dotierung durch geeignete Prozessoptimierung nutzbar gemacht werden sollte.

Übergangstiefe

Die bei den erhaltenen Profilformen gefundene Abhängigkeit der Flächenwiderstände von der ermittelten Übergangstiefe ist in der Abbildung 5.55 dargestellt. Dies erfolgte im Vergleich mit den in Abbildung 2.4 angegebenen Werten der geforderten Flächenwiderstände für die flachen Übergänge der SDEs der jeweiligen Technologiegenerationen laut der ITRS-Roadmap. Darüber hinaus sind die Werte mit eingezeichnet, die sich aus der Annahme einer idealen Eindiffusion von einer Monolage Phosphor ergeben haben und in Abbildung 3.2 dargestellt sind.

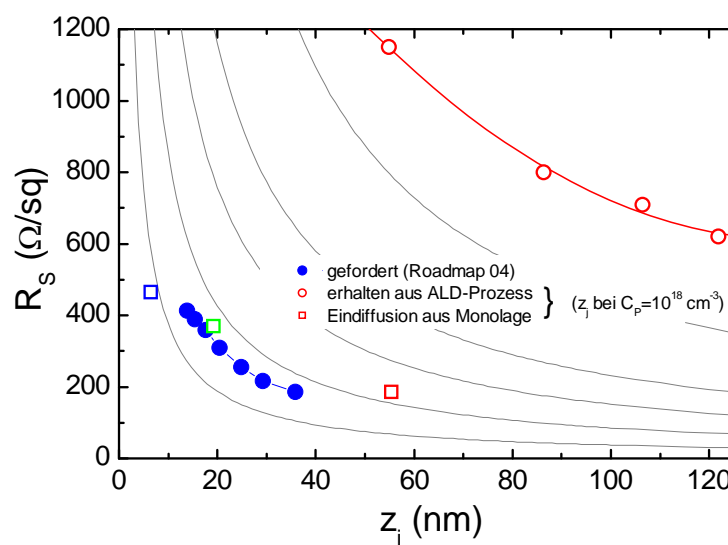


Abbildung 5.55: Zusammenhang zwischen Flächenwiderstand und Übergangstiefe von im zweistufigen Prozess dotierten Wafern im Vergleich zu den geforderten Werten laut Roadmap (siehe Abbildung 2.4) und errechneten Werten unter Annahme einer idealen Eindiffusion von einer Monolage Phosphor (siehe Abbildung 3.2).

Aus Abbildung 5.55 ist ersichtlich, dass das Dotierverfahren nach der Prozessvariante 2 mit den untersuchten Prozessparametern nicht den Anforderungen für flache Übergänge im Bereich der SDEs genügt. Für eine Erfüllung dieser Anforderungen sind offenbar weitere Prozessoptimierungen erforderlich, die eine noch geringere Übergangstiefe bei Erhöhung des Aktivierungsgrades der Dotierstoffmenge ermöglichen. Hierfür sollten insbesondere Flash-Annealing-Prozesse für den Eintreibeschritt erfolgen. Zusätzlich wird eine Erhöhung der Gesamtmenge des Phosphors durch eine optimierte oder auch mehrfache Belegung zu prüfen sein.

Diffusionskoeffizienten

Aus dem Parameter w der Gauß-Funktion kann durch Überführen dieser Funktion in die aus der Diffusionsgleichung bei Diffusion aus einer erschöpflichen Quelle abgeleitete Profilform der Konzentration,

$$C(z,t) = \frac{C_D}{\sqrt{\pi D t}} e^{-\frac{z^2}{4Dt}}, \quad (18)$$

der (effektive) Diffusionskoeffizient D_P für die Diffusion von Phosphor in Silicium errechnet werden:

$$D_P = \frac{w^2}{8t_{Dot}}. \quad (73)$$

Die Diffusionszeit t_{Dot} ist hierbei die Dauer des Diffusionsschrittes bei der hohen Eintreibetemperatur. Die so aus den Gaußprofilen ermittelten Diffusionskoeffizienten sind in der Tabelle 5.6 mit angegeben und in der Abbildung 5.56 in Abhängigkeit von der maximalen Eintreibetemperatur dargestellt. Zum Vergleich sind hierin die Temperaturverläufe der intrinsischen Diffusionskoeffizienten aus den Literaturwerten entsprechend Abbildung 3.4 hinterlegt.

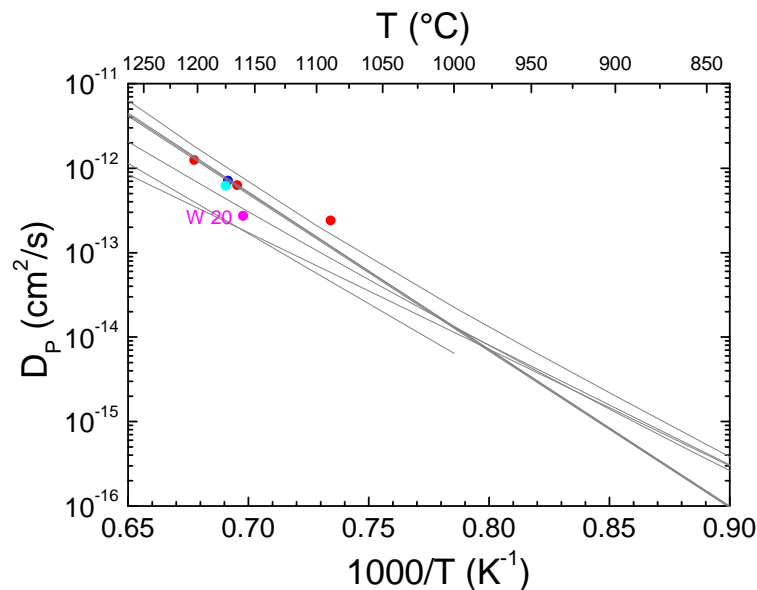


Abbildung 5.56: Vergleich der aus den Profilen ermittelten Diffusionskoeffizienten mit den intrinsischen aus Abbildung 3.4.

Aus dieser Darstellung wird deutlich, dass die Diffusionen durch die Annahme von intrinsischen Diffusionskoeffizienten gut beschrieben werden können. Eine offensichtliche Verstärkung der Diffusion durch transiente Effekte scheint nicht aufgetreten zu sein. Jedoch kann der höhere Wert des Diffusionskoeffizienten bei der geringeren Temperatur auf eine Verstärkung durch extrinsische Effekte hindeuten, wie auch ein Vergleich mit der intrinsischen Ladungsträgerkonzentration bei dieser Temperatur (siehe Abbildung 3.3) zeigt. Zudem könnten hier, aufgrund einer noch höheren Oberflächenkonzentration und der noch größeren Nähe des Profils zur Oberfläche und damit in deren verstärktem Einfluss, oberflächenbeeinflusste Transportvorgänge eine relativ größere Rolle gespielt haben.

Für genaue Schlussfolgerungen ist jedoch auch die exakte Kenntnis und Einbeziehung des Temperaturverlaufes während der Diffusion erforderlich. Dies zeigt sich auch anhand des scheinbar niedrigeren Diffusionskoeffizienten von Wafer 20. Da dieser Wafer jedoch mit längerer Eintreibzeit prozessiert wurde, ist hierfür die einfache Annahme der Maximaltemperatur für die Beschreibung der Diffusion nicht mehr ausreichend, da deren Einfluss auf die effektive Diffusionstemperatur geringer gewesen ist (siehe Beispiel für den Temperaturverlauf in Abbildung 5.12). So müsste hier von einer geringeren effektiven Temperatur ausgegangen werden.

5.4.5.4 Untersuchungen des Pile-up

In den Konzentrationsprofilen der Waferproben, bei denen keine Flußsäurebehandlung vor den SIMS-Analysen durchgeführt wurde, ist eine signifikant erhöhte Phosphorkonzentration nahe der Oberfläche in den Profilen sichtbar (siehe Abbildung 5.52). Ein Ausschnitt eines Profils für diesen oberflächennahen Bereich von Wafer 10 ist in der Abbildung 5.57 angegeben. Die Ergebnisse der SIMS-Messungen für Sauerstoff und Silicium sind hier ebenfalls eingefügt.

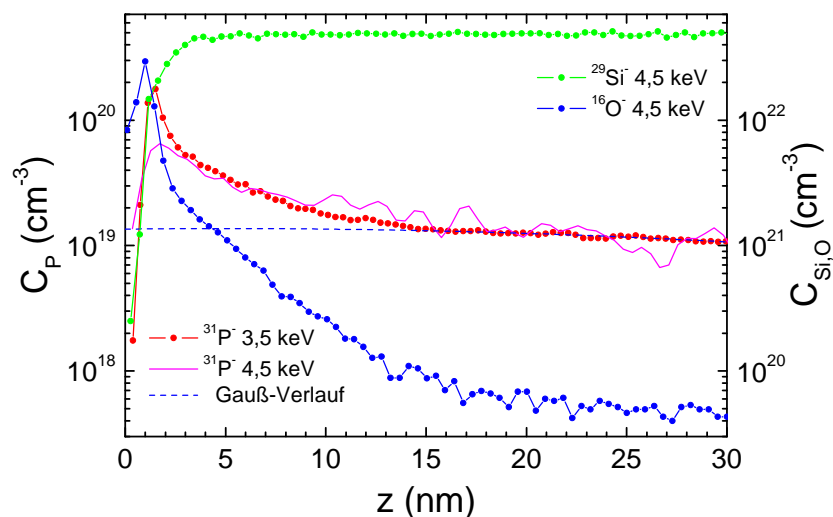


Abbildung 5.57: SIMS-Tiefenprofile von $^{31}\text{P}^-$, $^{16}\text{O}^-$ und $^{29}\text{Si}^-$ nahe der Oberfläche eines Siliciumwafers nach dem Zwei-Stufen-Dotierverfahren bei $T_{\max} = 1165^\circ\text{C}$ und $t_{\text{Dot}} = 10\text{ s}$. Die Gauß-Näherung des Phosphorprofils ist gestrichelt dargestellt.

Die Profilierung von Interfaces und Oberflächen durch SIMS stellt nach wie vor eine Herausforderung dar, selbst bei geringsten Primärionenenergien [de Chambost '04]. Qualitative

Ergebnisse für diesen Bereich können jedoch von den Messungen abgeleitet werden. Zwei unterschiedlich stark zum Silicium hin abfallende Flanken erscheinen in den Profilen von Phosphor und Sauerstoff. Die Lage des Maximums des Phosphor-Peaks liegt innerhalb der steilen abfallenden Flanke des Sauerstoff-Peaks, die die Position der Grenzfläche zwischen Oxid und Silicium anzeigt. In dieser oberflächennahen oder Interface-Region, deren Ausdehnung auch anhand des ^{29}Si -Gradienten abgeschätzt werden kann, ist nicht zu erwarten, dass die Messpunkte quantitativ korrekt sind, da sich sowohl die Sekundärionen-Ausbeute als auch die Sputterrate im Oxid und in der Grenzfläche vom Matrixwert unterscheiden. Jedoch können nicht alle Unterschiede des SIMS-Profiles zur Gauß-Approximation in diesem Bereich als Artefakte der SIMS-Messung interpretiert werden. Wie zuvor beschrieben, gibt es Hinweise für die Annahme, dass tatsächlich eine erhöhte Phosphorkonzentration an der Oberfläche (am Interface) nach dem Dotierprozess noch vorlag. Es ist bekannt, dass ein exponentieller Abfall der Sauerstoffkonzentration bei der SIMS-Profilierung von oxidbedecktem Silicium durch Ionenstrahlmischen von Sauerstoff in die Silicium-Matrix hinein auftritt. Das Phosphorprofil zeigt einen ähnlichen Verlauf, der somit durch den gleichen Effekt hervorgerufen worden sein kann, wenn eine Phosphorquelle vor der SIMS-Messung an der Oberfläche noch vorhanden war.

Eine andere mögliche Erklärung für den Abfall der Phosphorkonzentration kann eine Segregation von Phosphor aus Interface-Haftstellen (traps) während des Eintreibschritts sein. Neben der freien Siliciumoberfläche selbst stellt auch die SiO_2/Si -Grenzfläche eine effiziente Senke für Dopanten dar. (Damit wäre auch bei geringer Oxidbedeckung eine Belegung der „Oberfläche“ (d.h. hiermit der Grenzfläche) zu erklären.) Das Haftverhalten der Grenzfläche ist aus Pile-up- und Dose-loss-Studien bekannt [Sato '97; Vuong '00]. Außerdem wurde beobachtet, dass der Trapping-Prozess an der SiO_2/Si -Grenzfläche bei hohen Ausheiztemperaturen reversibel ist [Griffin '95]. Solch ein Lösen aus Haftstellen, unterstützt durch den Sauerstoff, der sich wie zuvor beschrieben als erforderlich herausstellte, sollte der normalen Diffusion vorausgegangen sein. Da bei solchen Segregationsprozessen ein Gleichgewicht zwischen Haften und Lösen angestrebt wird, ist auch nach dem kurzzeitigen Eintreibschritt noch Phosphor an der Oberfläche gebunden, der nicht zur aktiven Dotierung beiträgt.

Dies kann auch in Zusammenhang mit dem beobachteten Anstieg des Flächenwiderstands an Luft interpretiert werden. Ein Anhäufen von Phosphor (pile-up) während des Wachstums des natürlichen Oxids würde Dopanten in dem oberflächennahen Bereich deaktivieren, die vor der Oxidation bereits auf aktiven Gitterplätzen im Silicium waren. Der Phosphor-Pile-up und die beschleunigte Oxidation von an der Oberfläche hoch mit Phosphor dotierten Siliciumsubstraten wurde in verschiedenen Studien beschrieben [Kamiura '93; Sato '93].

Die Ergebnisse sind in Einklang mit den Dose-loss-Untersuchungen, die in der Literatur beschrieben wurden. Dort wird auch gezeigt, dass der Pile-up-Phosphor am Interface nicht aktiv ist und nicht als flacher Donator wirkt.

HF-Behandlung

Der Oberflächenpeak der Phosphorkonzentration war hingegen nicht ausgeprägt auf Wafern, bei denen vor der SIMS-Messung mittels HF-Dip das Oxid entfernt wurde. Die Abbildung 5.58 zeigt SIMS-Profile von Phosphor der Waferproben, die für 3 min in 1 % Fluorwasserstoffsäure geätzt wurden, im Vergleich zu einem Profil einer ungeätzten Probe. Die HF-Behandlungen wurden einige Tage vor den SIMS-Messungen ausgeführt, so dass bei Beginn der Messung auch ein natürliches Oxid vorgelegen hatte.

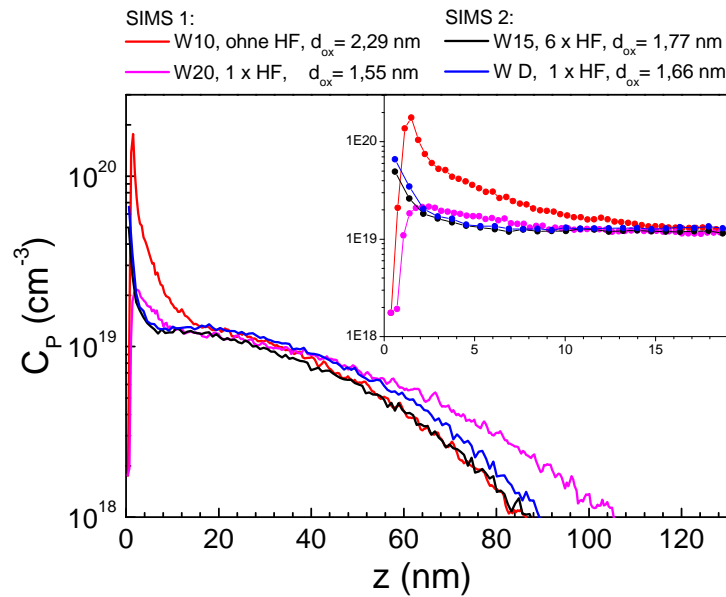


Abbildung 5.58: SIMS-Tiefenprofile von Phosphor in Silicium nach der Zwei-Stufen-Dotierung und unterschiedlicher HF-Behandlung vor den Analysen (SIMS 1: $E_{Cs} = 3,5$ keV, SIMS 2: $E_{Cs} = 0,5$ keV). Angegeben sind die Anzahl der HF-Behandlungen und die mittels Ellipsometer bestimmten Schichtdicken. Der Bereich nahe der Oberfläche mit erhöhter Phosphorkonzentration ist oben rechts vergrößert dargestellt.

Der Oberflächenphosphor ist offenbar mit der HF-Behandlung entfernt worden. Dies wurde auch in der Literatur beschrieben [Sato '93]. Dort wurde gefolgert, dass Oberflächenphosphor, der z.B. nach Hochtemperaturbehandlung implantierter Proben dort als Pile-up angehäuft ist (was somit zu dem so genannten Dose-loss führt), an der Oxidseite der Grenzfläche lokalisiert und damit durch die HF-Behandlung entfernbar ist.

Außerdem ist in der Abbildung 5.58 ein Vergleich von Profilen angegeben, die mit der geringen Cs^+ -Primärionenenergie von 0,5 keV aufgenommen werden konnten. Auch bei diesen Proben erfolgte vorher eine HF-Behandlung. Dabei ist zu erkennen, dass auch bei diesen Profilen ein geringer Phosphorpeak an der Oberfläche auftritt. Dieser ist jedoch schmaler, als der Peak der Probe, die mit 3,5 keV gemessen wurde (Wafer 20, mit HF-Behandlung), aufgrund einer geringeren Tiefe des transienten Bereiches, in dem die Empfindlichkeit für die Sekundärionen noch nicht konstant ist [Eto '05]. Da offenbar die Peakform von der Primärionenenergie abhängt und nicht von der Vorbehandlung, bedeutet dies, dass die Ausprägung dieses Peaks im Wesentlichen ein Artefakt der SIMS-Messung darstellt. Eine genaue Quantifizierung einer eventuell noch an der Oberfläche verbliebenen Phosphormenge und deren Lokalisation ist damit auch anhand der Ergebnisse der SIMS-Analysen mit den geringen Primärionenenergien nicht möglich.

Aus dem Vergleich der einfachen HF-Behandlung mit dem Ergebnis der Probe, die nach jeweils vorherigem Wachstum eines natürlichen Oxids mehrfach mit Flußsäure behandelt wurde, ist zu erkennen, dass nach den wiederholten HF-Behandlungen kein Unterschied in der Profilform nahe der Oberfläche aufgetreten ist. Die mehrfachen Behandlungen führen weder zu einer Veränderung des Oberflächenpeaks, was damit auch für dessen Herkunft aus der SIMS-Messung selbst spricht, noch kommt es durch die wiederholten natürlichen Oxidationen des Siliciums und dem anschließenden Entfernen der Oxidschichten zu einer Störung des Dotierprofils durch An- oder Abreicherungen von Phosphor infolge der Oberflächenbehandlungen.

5.4.6 Untersuchungen der Belegung

In der Abbildung 5.59 ist das Dotierprofil eines dotierten Wafers zweien gegenübergestellt, bei denen kein Eintreiben und somit nur die Belegungsschritte erfolgten. Wenn dafür vor der Belegung noch ein Ausheizen in der Dotiergasatmosphäre durchgeführt wurde, ist ein hoher Flächenwiderstand messbar (ca. 9000 Ω/sq für Wafer 9), erfolgte ausschließlich die Belegung, war kein Flächenwiderstand messbar.

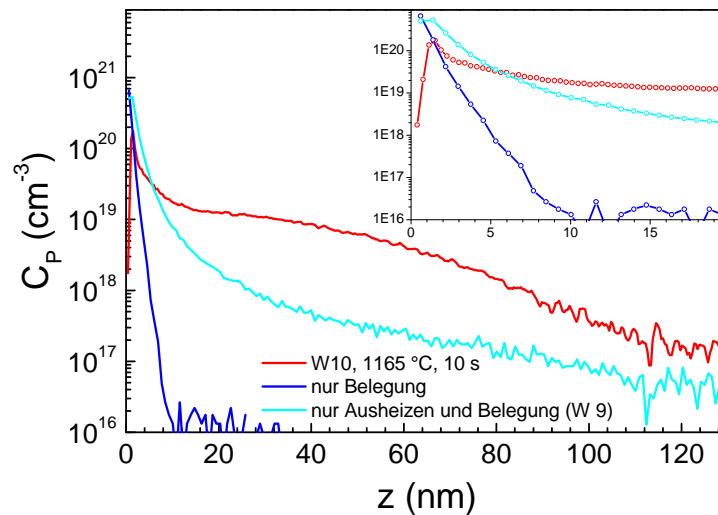


Abbildung 5.59: Vergleich des Phosphorprofils in Silicium eines mittels Prozessvariante 2 dotierten Wafers mit denen von Wafern, womit nur eine Belegung bzw. eine Belegung mit vorherigem Ausheizen in Dotiergasatmosphäre durchgeführt wurde. Das eingefügte Diagramm zeigt vergrößert den Bereich bis zu einer Tiefe von 20 nm.

Der Vergleich zeigt nun, dass nach der Belegung eine sehr hohe Phosphorkonzentration an der Oberfläche vorlag, bei vorherigem Ausheizen jedoch auch zusätzlich bereits Phosphor in das Silicium eindiffundiert war und so bereits zu dem zwar hohen, doch messbaren Flächenwiderstand führte. Aus dem Vergleich der Flächenwiderstände von solchen unterschiedlich vorbelegten Wafern *nach* dem Eintreibeschritt (z.B. Wafer 2 oder 3 mit Wafer 10) folgt jedoch, dass trotz des bereits beim Ausheizen eindiffundierten Anteils diese Wafer einen deutlich niedrigeren Flächenwiderstand aufwiesen als die Wafer, bei denen nur die Vorbelegung erfolgte.

Das kann nun folgende Gründe haben. Die Waferoberfläche kann nach dem Ausheizen nicht mehr so effektiv und mit der notwendigen Phosphordosis belegt werden und somit kann auch weniger Phosphor in das Silicium im Eintreibeschritt eindiffundieren. Dann müsste aber die eigentliche Belegung der Oberfläche in den obigen SIMS-Profilen nicht sichtbar oder vor den SIMS-Messungen von der Oberfläche verloren gegangen sein, da der ausgeheizte Wafer eine deutlich höhere Phosphorkonzentration aufweist. Oder die ausgeheizte und offenbar durch Ätzprozesse gestörte Siliciumoberfläche vermindert eine effektive Eindiffusion von Phosphor durch zusätzliche Oberflächenhaftstellen oder fördert eine Desorption von Phosphor gegenüber der Eindiffusion trotz der Sauerstoffatmosphäre. Aus den SIMS-Messungen allein lassen sich jedoch keine quantitativen Aussagen zur Belegung auf der Oberfläche vor der Eindiffusion ableiten.

Aufgrund dieser besonderen Schwierigkeit einer quantitativen Auswertung von SIMS-Profilen aus Tiefenbereichen in der unmittelbaren Nähe von Interfaces oder Oberflächen wurde versucht, den Phosphor, der nach der Belegung und vor dem Eintreiben an der Ober-

fläche vorliegen muss, mit AES nachzuweisen und dessen Menge zu ermitteln. Außerdem wurde dabei auch der Einfluss einer HF-Behandlung auf die Oberflächenbelegung untersucht.

In der Abbildung sind die AES-Spektren zweier Wafer dargestellt, bei denen nur die Phosphorbelegung erfolgte. Bei einem dieser Wafer wurde zusätzlich vor der AES-Messung eine HF-Behandlung durchgeführt. Man kann in dieser Darstellung gut die Veränderung einzelner Peaks erkennen.

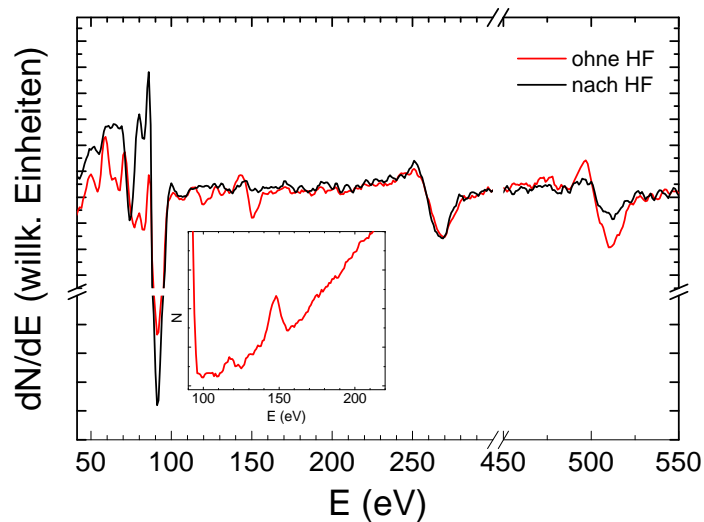


Abbildung 5.60: AES-Spektren zweier Wafer, bei denen nur die Belegung aber kein Eintreiben erfolgte. Der Unterschied ist eine HF-Behandlung des einen Wafers vor der Messung, das eingefügte Diagramm zeigt einen Teil des Rohspektrums des nicht mit HF behandelten Wafers vor der Ableitung.

Die Peaklagen sind in der folgenden Abbildung mit den jeweiligen Elementbezeichnungen markiert.

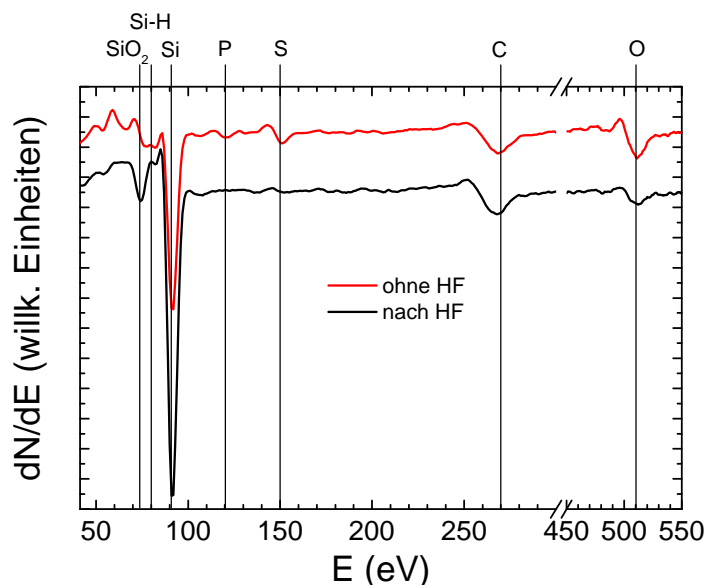


Abbildung 5.61: Vergleich der AES-Spektren aus Abbildung 5.60 mit eingezeichneten Peaklagen.

Aus diesen Abbildungen ist abzuleiten, dass Phosphor nach der Belegung an der Oberfläche nachzuweisen ist, der zugehörige Peak jedoch sehr schwach ausgeprägt ist, so dass eine

quantitative Messung auch hierfür nicht aussagekräftig ist. Nach einer HF-Behandlung ist dieser Peak nicht mehr nachzuweisen. Offenbar ist die Belegung dadurch entfernt worden. Die HF-Behandlung beeinflusst außerdem den Sauerstoff- und Siliciumanteil. Während der von Sauerstoff dadurch geringer wird, steigt der von Silicium an, was zu erwarten ist. Kontaminationen der Oberfläche mit Kohlenstoff werden durch die HF-Behandlung kaum beeinflusst, diese erfolgen wahrscheinlich wieder unmittelbar danach, während des Transports zur AES-Analyse. Darüber hinaus wurden Peaks gefunden, die auf Kontaminationen mit Schwefel hindeuten. Dieser ist vor der HF-Behandlung deutlich ausgeprägt und auch danach noch in geringem Ausmaß nachweisbar. Es sind organische Verbindungen in Reinraum-atmosphären bekannt, die als Ursache von Schwefelkontaminationen in Frage kommen [Habuka '03; Den '06].

5.4.7 Oberflächeneigenschaften der dotierten Wafer

Zur Überprüfung des Einflusses der Dotierprozesse auf die Oberflächeneigenschaften wurden Ellipsometermessungen der dotierten Wafer vorgenommen. Als Referenzspektren dienen dabei die genau analysierten der gereinigten Wafer (siehe Kapitel 5.1).

Bei den mit vorherigem Ausheizen prozessierten Wafern zeigten sich erhebliche Abweichungen vom Referenzspektrum. In der Abbildung 5.62 ist ein solches Beispiel angegeben (Wafer 3, siehe Tabelle 5.4). Hierin ist nur das Δ -Spektrum dargestellt, in dem sich die deutlichsten Abweichungen zeigten.

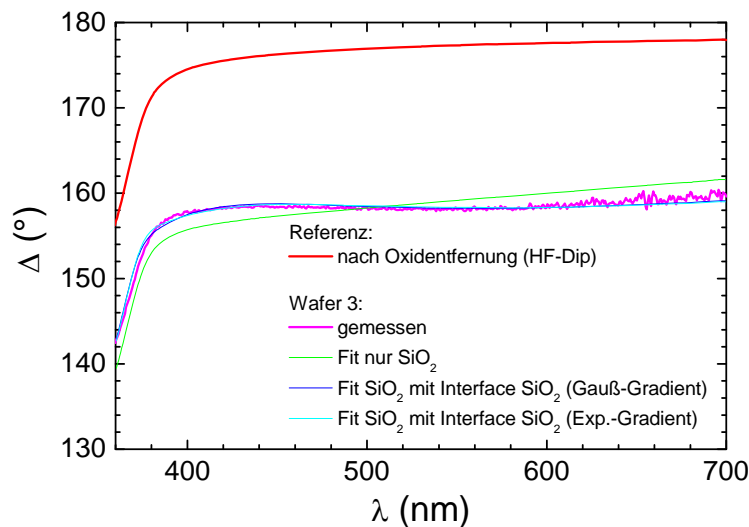


Abbildung 5.62: Beispiel für ein ellipsometrisches Δ -Spektrum von Wafer 3 (ca. Wafermitte) im Vergleich zum Referenzspektrum und Vergleich der jeweils günstigsten Approximation dreier Modelle.

Man kann in dieser Abbildung gut erkennen, dass die Abweichungen von einer HF-gereinigten oder nur mit Oxid bedeckten Siliciumoberfläche so groß sind, dass die Annahme des einfachen Modells (Einschichtmodell, wie in Abbildung 4.9 und folgender Abbildung angegeben) keine zufriedenstellende Repräsentation der tatsächlichen, das polarisierte Licht beeinflussenden Struktur darstellt. Die mit dem einfachen Modell errechneten ähnlichsten Spektren weisen eine zu große Differenz zu den gemessenen auf, wie der Vergleich der grünen Kurve mit der gemessenen zeigt.

Daraufhin wurden komplexere Schichtmodelle untersucht, von denen in der Abbildung 5.63 drei im Vergleich zu dem Einschichtmodell schematisch dargestellt sind.

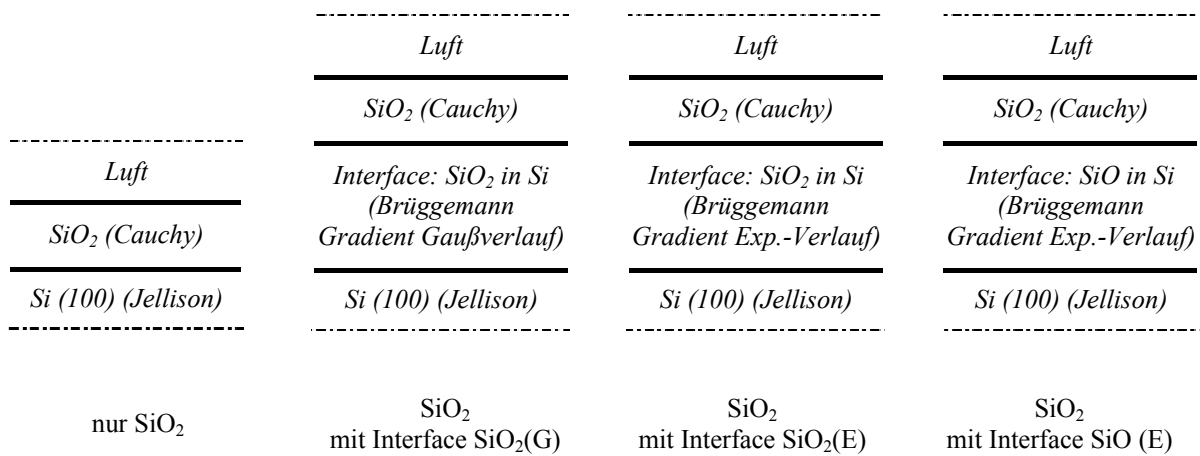


Abbildung 5.63: Verwendete Modelle für die Auswertung der ellipsometrischen Spektren.

Eine einfache Annahme von veränderten optischen Eigenschaften der Schicht des Einschichtmodells, z.B. über die Variation der wellenlängenabhängigen Brechzahlen der Oxidschicht selbst, führte zu keinem sinnvollen Ergebnis. Nur unter der Annahme von Interface-Schichten, die eine tiefenabhängige Einmischung eines Materials in ein Substratmaterial repräsentieren (hier: Brüggemann effektives Medium mit Gradient) wurde eine zufriedenstellende Modellierung möglich. Die Materialien und Gradienten wurden so gewählt, wie sie gemäß der Prozessierung zu erwarten wären. In der Abbildung 5.62 sind die Ergebnisse zweier Schichtmodelle mit angegeben. Hierbei wurde ein Interface aus einer Einmischung von SiO₂ in Si angenommen, mit jeweils unterschiedlicher Form der Gradienten. Nach Variation über die Dicke der beteiligten Schichten und der Mischungsparameter ergibt sich jeweils ein Modell, das die Ergebnisse der ellipsometrischen Messungen sehr gut repräsentiert. Die damit ermittelten Interface-Schichten sind in der Abbildung 5.64 angegeben, und die relevanten Modellparameter sind in der Tabelle 5.7 zusammengefasst.

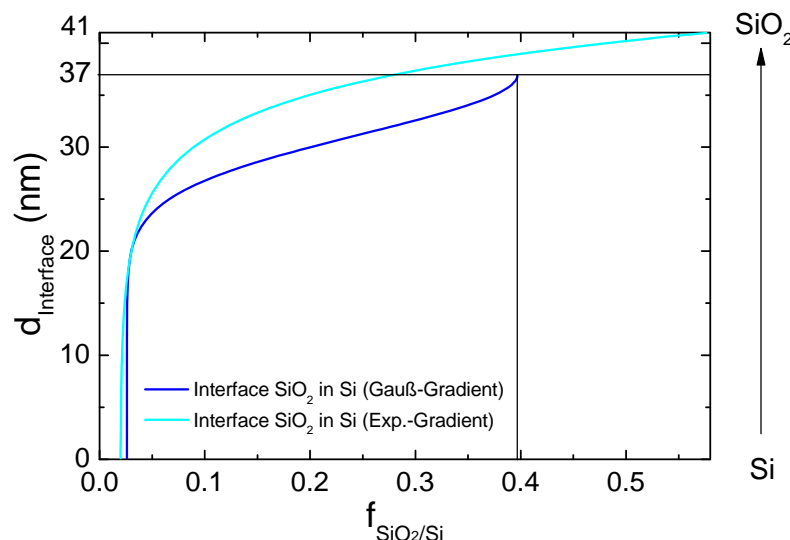


Abbildung 5.64: Darstellung der ermittelten Gradienten der jeweiligen Interface-Modelle für den Fit aus Abbildung 5.62.

Aus der Darstellung der Gradienten erschließt sich gut deren Struktur. Der Wert $d_{Interface}$ zeigt die Dicke der ermittelten Zwischenschicht an, sie ist mit 37 - 41 nm für beide Modelle in etwa gleich. Mit f ist der Anteil des eingemischten Mediums (hier: SiO₂) in das Silicium bezeichnet. Dessen maximaler und minimaler Wert sind in der Tabelle mit angegeben. Außerdem ist dort der Fitparameter σ_G als Maß für die Breite des Gradientenverlaufs mit angeführt.

Tabelle 5.7: Übersicht über die aus den Approximationen ermittelten Modellparameter für die in Abbildung 5.62 dargestellten Spektren.

Wafer	Modell	MSE	d_{Oxid} (nm)	$d_{Interface}$ (nm)	f_{min}	f_{max}	σ_G
Referenz (nach HF-Dip)	nur SiO ₂	0,2259	0,58	-	-	-	-
Wafer 3	nur SiO ₂	1,800	6,67	-	-	-	-
	SiO ₂ mit Interface SiO ₂ (G)	0,4678	0	37,00	0,26	0,397	3,248
	SiO ₂ mit Interface SiO ₂ (E)	0,4729	0	41,08	0,02	0,587	3,888

Der MSE-Wert in der Tabelle, das Maß für die Übereinstimmung von gemessenem und berechnetem Verlauf, zeigt die große Abweichung für das Einsichtmodell, während die Qualität der Approximation für beiden verschiedenen Gradienten in etwa gleich und bedeutend besser ist. Bei der Modellierung mit der Interfaceschicht ergibt sich, dass keine weitere reine Oxidschicht auf dem Wafer vorliegen würde, das ‚Interface‘ somit die Oberflächenschicht darstellen würde.

Aus dem Spektrum lässt sich somit ablesen, dass der Prozess von Wafer 3 zu einem erheblichen Maße zu einer Störung der Siliciumoberfläche geführt hat, und nur unter der Annahme einer Interfaceschicht aus SiO₂ in Si, die auch als Rauigkeit interpretiert werden kann, wird das gemessene ellipsometrische Spektrum einigermaßen plausibel modellierbar.

Noch deutlicher sind die Abweichungen von einer ungestörten Oberfläche bei Wafer 1, der nach Prozessvariante 1 prozessiert wurde. Spektren, die an verschiedenen Stellen auf der Waferoberfläche aufgenommen wurden, sind in der Abbildung 5.65 zusammengestellt. (Hier ist auch nur das Δ -Spektrum angegeben, obwohl bei diesem Wafer auch Ψ -Spektrum deutliche Abweichungen vom ungestörten zeigte.) Bei diesem Wafer waren bereits mit bloßem Auge Veränderungen auf der Oberfläche auszumachen. Es war ein streifenförmiges Muster aus grauen und dunkleren Bereichen zu erkennen. In allen diesen Bereichen war der Lichtspot des Ellipsometers aufgrund von Streulicht sichtbar, bei HF-gereinigten oder Wafers mit nur dünner Oxidschicht hingegen nicht. Etwas Abseits der Mitte war in einem kreisförmigen dunklen Bereich in dem Streifenmuster ein heller Fleck auszumachen. Auch der Waferrand war dunkler und teilweise etwas verfärbt, hier waren bereits Versetzungsmuster in der polierten Oberfläche zu erkennen. Die Spektren der verschiedenen Stellen sind in der Abbildung markiert.

Ein Vergleich mit den Topographieabbildungen der Oxidationsversuche und der Verteilung des Flächenwiderstands dieses Wafers legt nahe, dass die dort erkennbaren Strukturen, den zuvor beschriebenen entsprechen, das Streifenmuster gibt die Lampenanordnung wieder und der Fleck liegt im Bereich der gefundenen Flächenwiderstandserhöhung.

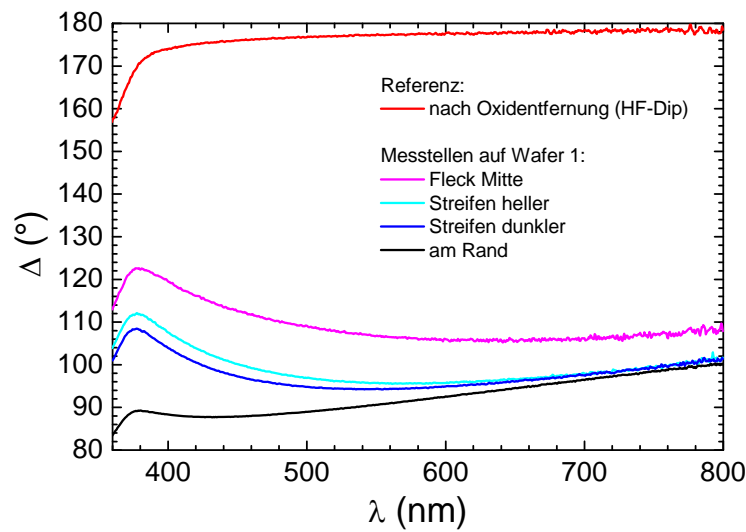


Abbildung 5.65: Delta-Spektrum von Wafer 1 gemessen an verschiedenen Stellen im Vergleich zum Spektrum eines HF-gereinigten Wafers.

Eine sinnvolle Modellierung der Oberflächenstruktur für eine Beschreibung der ellipsometrischen Spektren ist hierfür sehr schwierig. Beispiele dafür sind in der Abbildung 5.66 und der Tabelle 5.8 angegeben.

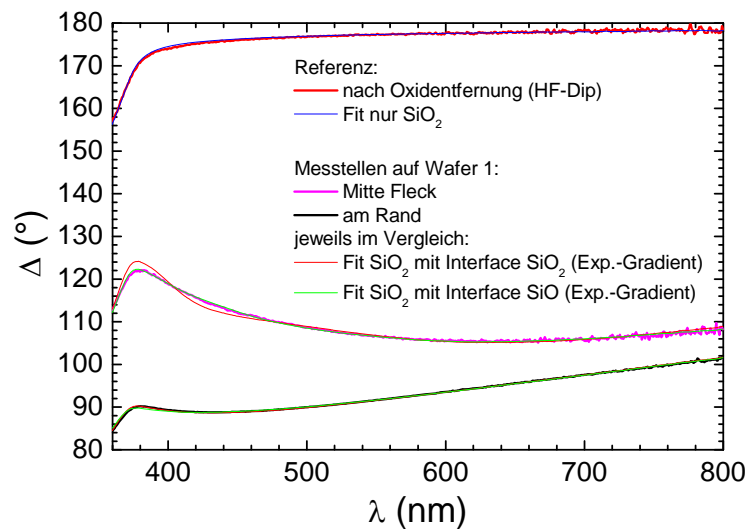


Abbildung 5.66 Vergleich ausgewählter Δ -Spektren von Wafer 1 mit den berechneten von günstigsten Modellen.

Die Spektren lassen sich nur durch eine Interfaceschicht von großer Ausdehnung nachvollziehen, wobei hier die Annahme einer Einmischung von SiO ein etwas besseres Modell darstellt. Die insbesondere am Rand scheinbar hohen Oxidschichtdicken erscheinen hingegen wenig plausibel. Wahrscheinlich haben die beobachteten Versetzungen zusätzliche, nicht mit gewöhnlichen Modellen beschreibbare Einflüsse auf das Spektrum.

Tabelle 5.8: Übersicht über die aus den Approximationen ermittelten Modellparameter für die in Abbildung 5.66 dargestellten Spektren.

Messposition	Modell	MSE	d_{Oxid} (nm)	$d_{Interface}$ (nm)	f_{min}	f_{max}	σ_G
Fleck bei Mitte	SiO ₂ mit Interface SiO ₂ (E)	0,5794	5,4	72,74	0,014	0,594	1,539
	SiO ₂ mit Interface SiO (E)	0,2301	8,36	143,39	0,002	0,635	2,640
Rand	SiO ₂ mit Interface SiO ₂ (E)	0,1968	21,44	98,19	0	0,759	3,527
	SiO ₂ mit Interface SiO (E)	0,1744	24,65	145,14	0	0,817	5,268

Aus den Ergebnissen wird deutlich, dass durch die für diese Wafer eingesetzten Dotierprozesse die Oberfläche deutlich aufgeraut wurde. Dies war offenbar die Folge eines Ätzens der Oberfläche in der Ausheizatmosphäre. Eine einfache Analyse von ellipsometrischen Spektren gibt Hinweise auf die Art und Größenordnung der Störung der Oberfläche. Für eine genaue quantitative Ermittlung der Oberflächenveränderungen sind jedoch zusätzliche vergleichende Messungen, wie mittels Atomkraftmikroskop erforderlich.

Für weitere dotierte Wafer nach der Prozessvariante 2 sind ausgewählte Ergebnisse der ellipsometrischen Messungen und deren Auswertungen in der Abbildung 5.67 und in der Tabelle 5.9 zusammengefasst.

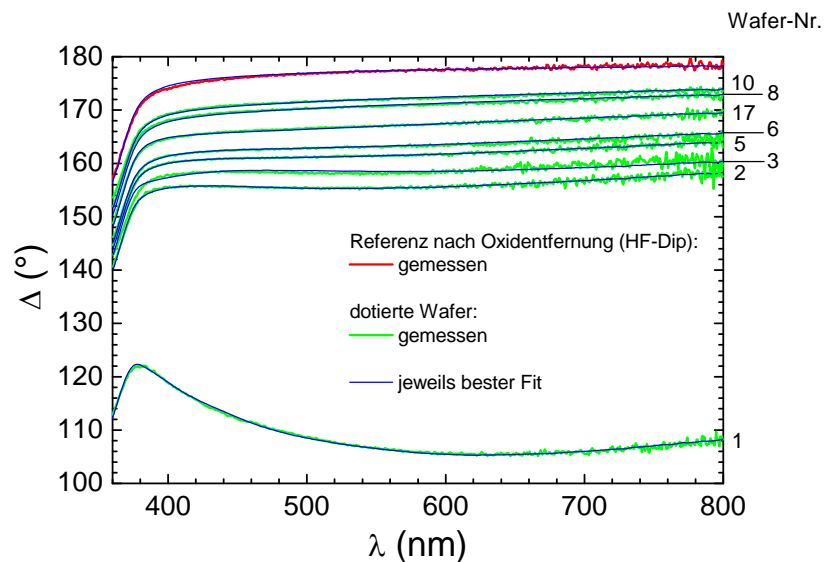


Abbildung 5.67: Vergleich ausgewählter Δ -Spektren verschiedener dotierter Wafer (gemessen in der Wafermitte) mit dem Referenzspektrum. Blau dargestellt: berechnete Spektren aus günstigsten Modellen.

Tabelle 5.9: Übersicht über die aus den Approximationen der ellipsometrischen Spektren ermittelten Modellparameter für verschiedene dotierte Wafer (gleiche Wafernummern wie in Tabelle 5.4).

Wafer	Modell	MSE	d_{Oxid} (nm)	$d_{Interface}$ (nm)	f_{min}	f_{max}	σ_G
2	SiO ₂ mit Interface SiO ₂ (E)	0,2399	3,88	35,45	0,015	0,301	3,024
4	SiO ₂ mit Interface SiO ₂ (E)	0,1999	2,78	0	-	-	-
5	SiO ₂ mit Interface SiO ₂ (E)	0,1695	3,30	35,15	0,009	0,228	3,577
6	SiO ₂ mit Interface SiO ₂ (E)	0,1662	2,97	34,69	0,008	0,234	4,156
8	nur SiO ₂	0,2442	2,69	-	-	-	-
10	nur SiO ₂	0,2320	2,29	-	-	-	-
11	nur SiO ₂	0,2369	2,08	-	-	-	-
12	nur SiO ₂	0,2443	2,47	-	-	-	-
17	SiO ₂ mit Interface Luft (E)	0,1943	2,27	13,66	0,022	0,563	8,990
20	nur SiO ₂	0,2118	1,55	-	-	-	-

Aus diesen Ergebnissen lässt sich ableiten, dass die deutlichen Abweichungen von der HF-gereinigten oder nur mit Oxid bedeckten Oberfläche, die sich nur unter der zusätzlichen Annahme von Interfaceschichten¹ modellieren lassen, *immer* als Folge von Prozessen auftraten, die ein Ausheizen oberhalb von ca. 800 °C ohne Sauerstoffzugabe beinhalteten. Hingegen führten die Prozesse, bei denen nur eine Belegung bei moderaten Temperaturen ohne Sauerstoffzugabe durchgeführt wurde und ein anschließendes Eintreiben in sauerstoffhaltiger Atmosphäre erfolgte, zu Ergebnissen, die durch simple Beschreibung mittels einer einfachen Oxidschicht auf einem ungestörten Siliciumsubstrat zu modellieren sind, genau so, wie sie sich auch für Siliciumoberflächen mit chemischem Oxid (z.B. nach Caro-Behandlung) oder nach einfacher thermischer Oxidation (siehe Wafer 8) ergeben. Als Beispiel ist in der Tabelle 5.9 für die Modellierung des Spektrums von Wafer 4 angegeben, dass selbst bei der Annahme einer Interfaceschicht im Modell, diese für die beste Näherung des Modells die Dicke Null haben würde. Damit sollte das gewählte Dotierverfahren gemäß der Prozessvariante 2 eine zerstörungsfreie und strukturerhaltende Methode zu Dotierung von Silicium darstellen.

¹ Verschiedene Interfacemodelle wurden zusätzlich getestet, wobei die beschriebenen EMA-Modelle mit Gradienten die besten Ergebnisse erbrachten. Geringfügige Verbesserungen wurden hierbei im Einzelfall noch erzielt, wenn statt SiO₂ oder SiO als eingemischtes Material auch Luft angenommen wurde, was insbesondere für Modellierungen von Rauigkeiten verwendet werden kann. Aufgrund der geringen Unterschiede der Ergebnisse dieser Modelle untereinander kann jedoch hieraus allein keine eindeutige Aussage zur genauen Struktur der Interfaceschichten getroffen werden.

Zusammenhang von Dotierung und Oxidwachstum

Da sich die ellipsometrischen Spektren der mittels Prozessvariante 2 dotierten Wafer durch das einfache Modell gut beschreiben lassen, sind hierfür einfache vergleichende Messungen für die Einflüsse von unterschiedlichen Prozessparametern und auch ortabhängige Messungen von Schichtdickentopographien möglich.

In der Abbildung 5.68 sind die Δ -Spektren von vier dotierten Wafern dargestellt. Aus dem Vergleich der Wafer 10 - 12 ist zu erkennen (siehe auch Tabelle 5.4 Tabelle 5.9), dass mit höherer Eintreibtemperatur auch eine höhere Schichtdicke einhergeht. Nach einer der Dotierung folgenden HF-Behandlung, wie sie vor der Messung von Wafer 20 erfolgte, ist hingegen nicht wieder der Zustand der undotierten, oxidfreien Oberfläche erreicht worden. Es wurde hierfür noch eine Schichtdicke von 1,55 nm ermittelt.

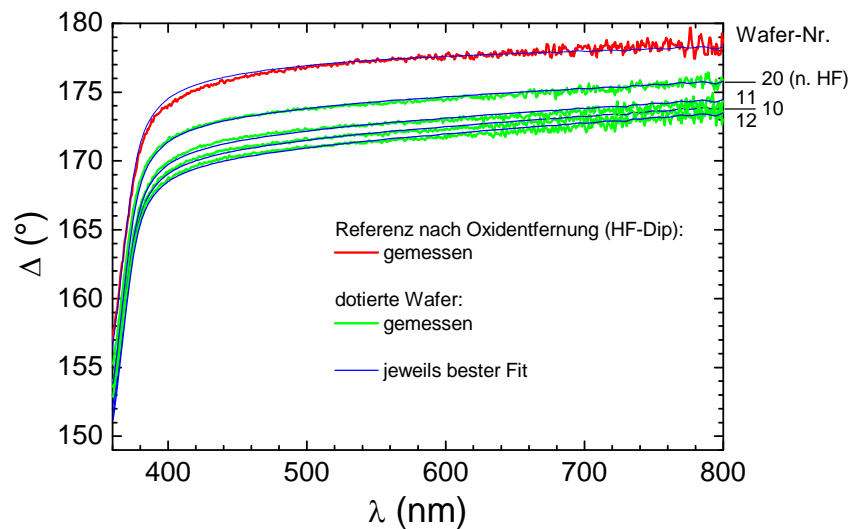


Abbildung 5.68: Vergleich der Δ -Spektren von dotierten Wafer (gemessen in der Wafermitte) mit und ohne HF-Behandlung nach dem Dotierprozess mit dem Referenzspektrum. Blau dargestellt: berechnete Spektren aus günstigsten Modellen.

Bei der Analyse der Spektren fällt weiterhin auf, dass hierin keine Einflüsse einer Phosphorbelegung oder eines Phosphor-Pile-up deutlich werden. Bei ausschließlich erfolgter Belegung wurde eine Schichtdicke von 1,42 nm (MSE = 0,2107) gemessen. Auch die Annahme einer dünnen amorphen Si-Interfaceschicht, wie im Kapitel zur Eingangsreinigung diskutiert, ergab bei diesen Spektren keine Verbesserung des Modells. Eindeutige Aussagen sind daher mit diesen Messungen nicht möglich, da solche Einflüsse im betrachteten Wellenlängenbereich zu gering sind oder sich kaum von denen einer angenommenen Oxidschicht unterscheiden.

In der Abbildung 5.69 ist ein Beispiel für Verläufe der mittels Ellipsometer unter Verwendung des einfachen Schichtmodells ermittelten Oxidschichtdicken für drei Wafer, die bei verschiedenen Eintreibtemperaturen prozessiert wurden, dargestellt. Zum Vergleich wurden in diese Abbildung auch die entsprechenden nach der Dotierung gemessenen Flächenwiderstände eingezeichnet. Hierbei erfolgte die Skalierung von hohem Widerstand (Skalenanfang) zu geringerem hin, um eine günstigere Vergleichbarkeit zu erzielen.

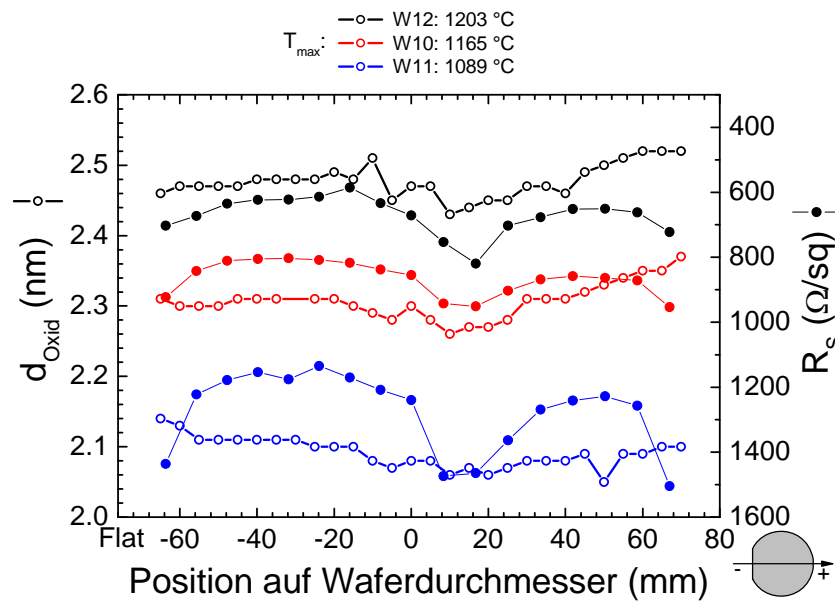


Abbildung 5.69: Verlauf der gemessenen Oxiddicke (offene Punkte) über dem Waferdurchmesser dreier bei unterschiedlicher Eintreibtemperatur dotierter Wafer im Vergleich zu den Flächenwiderständen.

Man erkennt in dieser Abbildung, dass die Flächenwiderstände eine größere Schwankung aufweisen, als die Oxiddicken. Besonders ausgeprägt ist dies bei der geringeren Eintreibtemperatur. Die typischen Erhöhungen der Flächenwiderstände etwas abseits von der Mitte sind nicht in dem Ausmaß im Oxiddickenverlauf abgebildet. Hier¹ ergibt sich für diesen Bereich eher eine geringere Schichtdicke in dem Bereich des höheren Widerstands, was auf geringere Prozesstemperaturen an dieser Stelle hindeutet. Deutliche Abweichungen treten auch an den Waferändern auf. Obwohl bei den Flächenwiderständen aller Wafer ein hier sehr deutlicher Anstieg zu verzeichnen ist, steigt auch die Schichtdicke in diesem Bereich eher an. Der Grund dafür ist nicht ganz klar. Es ist jedoch zu beachten, dass diese Resultate die Summe aller Prozesseinflüsse darstellen, insbesondere können sich neben der mittleren Prozesstemperatur die Aufheiz- und Abkühlrampen lokal sehr unterscheiden. Die Empfindlichkeit des Dotierens ist für diese Schwankungen offenbar höher als die der Oxidation, so dass der Einsatz des Dotierverfahrens als Monitorprozess für solche Schwankungen zu prüfen wäre.

¹ Im Gegensatz zu der bei den Temperaturkalibrierungen vorhandenen statischen Prozessatmosphäre wurde bei den Dotierversuchen mit Prozessvariante 2 mit stationären Sauerstofffluss gearbeitet.

6. Zusammenfassung und Ausblick

Die Dotierung von Silicium stellt einen grundlegenden Prozessschritt für die Fertigung von Halbleiterbauelementen dar. Ein verbreitetes Verfahren zur Dotierung ist dabei die Einbringung von Dotierstoff durch Ionenimplantation und ein nachfolgender Kurzzeit-Ausheizschritt für eine Aktivierung des Dotierstoffs im Silicium und eine Ausheilung von Strahlenschäden darin. Die Ausführung dieser Prozesse erfordert einen hohen technischen Aufwand. Prinzipielle Nachteile dieses Verfahrens, wie die Störung von oberflächennahen Bereichen des Substrats durch den Ionenbeschuss und dessen ausgeprägte Vorzugsrichtung rechtfertigen die Suche nach anderen Methoden. Als eine alternative Ausführung der Dotierung von Silicium wurde in der vorliegenden Arbeit ein Verfahren untersucht, bei dem eine Anlagerung von Dotierstoff auf der Siliciumoberfläche aus einer Gasphase heraus erfolgte und dessen Eindiffusion in oberflächennahe Bereiche in einem Kurzzeitprozess durchgeführt wurde. Als Dotierstoff wurde dabei Phosphor gewählt, der über die gasförmige Verbindung Phosphin bereitgestellt wurde. Die Durchführung dieses Verfahrens erfordert einen geringeren apparativen Aufwand als die Implantation, es treten keine Strahlenschäden im Silicium durch Ionenbeschuss auf und der Transport der Dotierstoffe aus der Gasphase erfolgt isotrop durch Diffusion an die Oberfläche.

Die Dotierexperimente wurden mit 150 mm-Siliciumwafern durchgeführt, mit denen eine nasschemische Standard-Eingangsreinigung mit Caro'scher Säure und anschließender Oxid-entfernung in verdünnter Fluorwasserstoffsäure erfolgte. Untersuchungen der Siliciumoberfläche nach dieser Vorbehandlung mittels spektroskopischer Ellipsometrie zeigten, dass die nasschemischen Oberflächenbehandlungen reproduzierbar durchgeführt werden konnten, jedoch keine ideale oxidfreie wasserstoffterminierte Oberfläche erhalten wurde. Ein Schichtwachstum der HF-gereinigten Wafer in der Reinraumluft wurde nachgewiesen. Die zu dotierenden Wafer wurden unmittelbar nach der Oxid-entfernung in eine Vakuumumgebung gebracht, so dass dieses Wachstum für den Dotierprozess keine Bedeutung hatte.

Die Dotierungen erfolgten in einer Kurzzeitprozessanlage mit konventioneller Lampenheizung unter Vakuumbedingungen. Die Möglichkeiten der Prozesskontrolle und -uniformität der Anlage wurden anhand von Testprozessen mit einem Thermoelement-Wafer und mittels Kurzzeitoxidationen in sauerstoffhaltiger Atmosphäre untersucht und bewertet.

Den zentralen Teil der vorliegenden Arbeit stellte die Untersuchung und weitere Entwicklung eines geeigneten Gasphasen-Dotierprozesses dar. Zunächst wurde ein einstufiger Gasphasenprozess untersucht und bewertet. Dabei erfolgte ein Aufheizen des Prozesswafers auf eine Temperatur, bei der eine Eindiffusion von Phosphor in das Silicium stattfand, während das Dotiergas Phosphin in der Prozessatmosphäre vorhanden war. Es zeigte sich jedoch, dass eine Phosphordotierung auf diese Weise ungünstig ist. Für ausreichende Phosphorkonzentrationen

im Silicium sind hohe Partialdrücke von Phosphin erforderlich. Dieses besitzt zwar einen hohen Haftkoeffizienten an Silicium mit wasserstoffterminierter Oberfläche und es erfolgt hier eine dissoziative Adsorption bei moderaten Temperaturen unterhalb von ca. 550 °C, jedoch ist die Desorptionsrate von Phosphor bei den Diffusionstemperaturen sehr hoch, so dass die Dotiereffizienz dieses Prozesses sehr gering ist. Entsprechende Versuche mit einem Partialdruck von Phosphin von 0,5 bis 1 mbar führten zu einer sehr schnellen Kontamination der Kammerinnenwände mit rotem Phosphor und erwiesen sich als schlecht reproduzierbar. Solche reinen Gasphasenprozesse sind schwierig zu kontrollieren, es können neben den überwiegend stattfindenden oberflächenkatalytischen Zerfallsprozessen des Phosphins hierbei auch Reaktionen bereits in der Gasphase auftreten, die schwer zu beeinflussen sind. Resultierende Kontaminationen sind für Halbleiterfertigungsprozesse nicht tolerierbar und führen zu Prozessinstabilitäten. Eine Minimierung der Kontaminationen ist durch Verminderung der Menge des eingesetzten Phosphins bei geringeren Prozesstemperaturen möglich; die Versuche zeigten jedoch, dass dann eine nennenswerte Dotierung des Siliciums mit Phosphor mit diesem Prozess nicht möglich ist.

Daher wurde ein alternativer Ansatz verfolgt. Es wurde eine Dotierung in zwei Stufen durchgeführt. Zunächst erfolgte eine Belegung der Siliciumoberfläche mit Phosphor durch dissoziative Adsorption von Phosphin bei einer Temperatur unterhalb der Diffusionstemperatur (maximal ca. 800 °C). Dieser Prozessschritt kann für eine effektive Oberflächenbelegung optimiert werden, wofür nur sehr geringe Dosen von Phosphin erforderlich sind. Es wurde 50 vpm Phosphin in Wasserstoff eingesetzt, mit einem Gasfluss von 100 sccm für maximal 60 s Belegungszeit. Die Belegung ist selbstbeschränkt und damit gut kontrollierbar. Die zweite Stufe war das Eintreiben des Phosphors von der Oberfläche in das Silicium in sauerstoffhaltiger Prozessatmosphäre. Eine weitere Dotierstoffzufuhr aus der Gasphase erfolgte hierbei nicht. Auf diese Weise war eine Eindiffusion von Phosphor ohne überwiegenden Verlust durch Desorption von der Oberfläche bei den erforderlichen Diffusionstemperaturen erreichbar. Eine Abscheidung einer Deckschicht aus Siliciumoxid, die bei vergleichbaren Prozessen zur Verhinderung der Desorption eingesetzt wird, konnte so entfallen.

Mit der Auswertung der Ergebnisse der Zweistufen-Prozesse wurde der Nachweis erbracht, dass auf diese Weise eine flache n-Dotierung von Silicium mit Phosphor bei geringem Einsatz von Phosphin möglich ist, wobei keine sichtbaren Kontaminationen der Kammer nach den durchgeführten Versuchsreihen gefunden wurden. Für das Eintreiben des Phosphors von der Oberfläche war ein geringer Sauerstoffdruck von 2,5 mbar ausreichend aber auch erforderlich. Günstigere Ergebnisse bezüglich Reproduzierbarkeit und Homogenität der Dotierung über die gesamte Waferfläche wurden für Prozesse erzielt, bei denen kein Ausheizen des Wafers in wasserstoffhaltiger Atmosphäre vor dem eigentlichen Dotierprozess erfolgte. Die nasschemisch gereinigte Siliciumoberfläche erwies sich somit als ausreichend für die dissoziative Adsorption von Phosphin.

Typische Dotierprozesse mit Maximaltemperaturen von 1090 bis 1200 °C bei Eintreibzeiten von 10 s ergaben pn-Übergangstiefen von 55 bis 125 nm und Flächenwiderstände von 1150 bis 620 Ω/sq . Die eindiffundierte Dosis des Phosphors lag dabei bei ca. $1 \cdot 10^{14} \text{ cm}^{-2}$. Die Analyse der aus SIMS-Messungen gewonnenen Dotierprofile ergab einen gaußförmigen Konzentrationsverlauf des eindiffundierten Phosphors, wobei auch nach dem Eintreibschritt noch ein Depot von Phosphor an der Oberfläche nachgewiesen werden konnte. Dieser Phosphor konnte durch eine Behandlung in verdünnter Flusssäure entfernt werden. Ein Vergleich von gemessenen Flächenwiderständen mit aus den Profilverläufen errechneten ergab, dass der eindiffundierte Phosphor elektrisch aktiv war, der Oberflächenphosphor hingegen nicht. Insgesamt war jedoch festzustellen, dass die Phosphordotierungen, die mit dem untersuchten Zweistufen-Verfahren für die Schaffung flacher pn-Übergänge nahe der Siliciumoberfläche

erzeugt wurden, den gegenwärtigen Anforderungen an Source-/Drain-Extensions von MOSFET-Bauelementen nicht genügten.

Oberflächenuntersuchungen der dotierten Wafer mittels spektroskopischer Ellipsometrie im Wellenlängenbereich von 350-800 nm zeigten, dass Wafer, bei denen während der Prozessfolge ein Aufheizen auf die Diffusionstemperaturen ohne Sauerstoffzufuhr erfolgte und für die große Inhomogenitäten der Dotierergebnisse gefunden wurden, eine deutlich gestörte Oberflächenstruktur aufwiesen, die als Rauigkeit interpretierbar ist, welche infolge eines Ätzens der Oberfläche durch Spuren von Sauerstoff auftrat. Hingegen ergaben ellipsometrische Messungen nach dem Dotierprozess aus Belegung bei moderater Temperatur in Phosphinatmosphäre und Eintreiben in Sauerstoff, dass die Oberflächenstruktur dieser Wafer durch ein einfaches Modell einer dünnen Schicht auf kristallinem Silicium beschrieben werden konnte, wie es auch bei gereinigten Wafern oder Wafern mit natürlicher Oxidschicht der Fall war. Eine Störung der Oberflächenstruktur durch diesen Dotierprozess wurde somit nicht gefunden. Eine eindeutige Bestimmung der Art der Schicht oder ein Nachweis eines Phosphordepots war mit diesen ellipsometrischen Messungen nicht möglich.

An den dotierten Wafern wurde darüber hinaus ein interessanter Effekt beobachtet. Während des ersten Kontakts mit der Reinraumluft zeigte sich ein deutlicher Anstieg des Flächenwiderstands. Es wurden mit einer neuen Messprozedur zeitabhängige Messungen während des Waferkontakts mit der Luft durchgeführt, wobei die vier Spitzen in Kontakt mit dem Wafer belassen wurden. Damit war eine sehr genaue Messung des Flächenwiderstandsanstiegs möglich. Dieser Anstieg wurde auch an dotierten Wafern nachgewiesen, nachdem eine Entfernung des natürlichen Oxids mit einer Flusssäurebehandlung erfolgt war und konnte mit dem Wachstum einer Schicht in Zusammenhang gebracht werden, das mit zeithabhängigen ellipsometrischen Messungen beobachtet wurde. Die Auswertungen der Anstiege ergaben Indikationen für einen Schicht-für-Schicht-Wachstumsprozess an der Waferoberfläche. Der Flächenwiderstandsanstieg von flach dotierten Schichten ist für die Bewertung der Dotierergebnisse von Bedeutung aber auch für Widerstandsprofilierungsverfahren, bei denen frisch präparierte Schlitze der Luft ausgesetzt sind. Die Methode der zeitabhängigen Flächenwiderstandsmessungen gestattet eine einfache Beobachtung von oberflächenbeeinflussten Widerstandsänderungen von flach dotierten Proben.

Aus den Ergebnissen der Untersuchungen lassen sich Ziele für weitere Arbeiten ableiten. Bei gleicher Belegung erscheint eine verbesserte Eindiffusion des Phosphors über eine effektivere Aktivierung des noch vorhandenen Depots an der Oberfläche durch Flash- oder Millisecond-Annealing-Verfahren viel versprechend. Damit sollten sich flachere pn-Übergänge mit höheren aktiven Phosphorkonzentrationen erzielen lassen. Durch Wiederholungen der Belegungs- und Eintreibeschritte ist die Möglichkeit einer Dosiserhöhung und -einstellung durch mehrfache Belegungen zu untersuchen. Dabei ist auch die Eignung des vorgestellten Prozesses für die Dotierung von dreidimensionalen Oberflächentopographien zu prüfen.

Weiterhin ist von Bedeutung, die Stabilität der Belegung bei Kontakt mit der Reinraumumgebung zu untersuchen. Sollte dabei die Belegung der Oberfläche mit Phosphor beständig und unempfindlich gegenüber Umgebungseinflüssen sein, was zu erwarten ist, ließen sich Belegungs- und Eintreibeschritte getrennt voneinander ausführen. Damit könnte der Oberflächenphosphor der vorbelegten Wafer in gewöhnlichen Kurzzeitprozessmodulen aktiviert werden, auch wenn diese nicht mit einer Phosphin-Gasversorgung ausgestattet sind. Außerdem eigneten sich diese vorbelegten Wafer dann für ein einfaches Temperatur-Monitoring in Annealing-Anlagen.

Unter Einsatz der Gase Arsenwasserstoff (Arsin) bzw. Antimonwasserstoff (Stibin), die ähnliche Eigenschaften einer dissoziativen Adsorption an Siliciumoberflächen besitzen wie das untersuchte Phosphin, kann die betrachtete Prozessfolge auch für eine Arsen bzw. Antimon-dotierung von Silicium eingesetzt und untersucht werden.

7. Danksagungen

An dieser Stelle möchte ich den Personen meinen Dank aussprechen, die mich bei der Durchführung meiner Arbeiten unterstützt haben. Besonderer Dank gilt Herrn Prof. Dr. Edmund P. Burte für die Anregung zur Beschäftigung mit dem interessanten Thema und für die Betreuung der Arbeit. Ich danke weiterhin Dr. Gerhard Bilger von der Universität Stuttgart und Dr. Gunther Lippert vom IHP Frankfurt (Oder) für die Ausführung von SIMS-Analysen und Dr. Wilfried Lerch von der Firma Mattson Thermal Products GmbH für die Bereitstellung von Thermoelement-Wafern. Mein Dank gilt auch den Studenten Denis Reso, Christian Domröse, Michael Burke und Thomas Löffler, die im Rahmen von Studienarbeiten bei der Bearbeitung der Aufgabenstellungen behilflich waren.

Schriftliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Insbesondere habe ich nicht die Hilfe einer kommerziellen Promotionsberatung in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

Magdeburg, den 4.12.2006

8. Liste der verwendeten Formelzeichen

Formelzeichen	Beschreibung	Einheit
a	Atomabstand im Siliciumkristall	$5,43 \cdot 10^{-10}$ m
A	Fitparameter der Gauß-Funktion (entspricht der Fläche unter der Funktion im Definitionsbereich)	cm^{-2}
A	Fitparameter für direkt-logarithmisches Gesetz	nm
B	Fitparameter für direkt-logarithmisches Gesetz	min^{-1}
C	Konzentration (allgemein)	cm^{-3}
C_D	Flächendichte (Dosis) von Phosphor im Silicium	cm^{-2}
C_{OF}	(Volumen-)Konzentration im Festkörper an (unterhalb) der Oberfläche	cm^{-3}
C_P	Phosphorkonzentration im Silicium (Volumenkonzentration)	cm^{-3}
C_{P_0}	Parameter der Gauß-Funktion (Verschiebung bezüglich C_P -Achse, entspricht der Phosphor-Untergrundkonzentration)	$7,5 \cdot 10^{13}$ cm^{-3}
C_{sat}	Sättigungskonzentration von gelöstem Dotierstoff im Gleichgewicht mit Präzipitaten	cm^{-3}
D	Diffusionskoeffizient (allgemein)	$\text{cm}^2 \text{s}^{-1}$
D^-	Diffusionskoeffizient von einfach negativ geladenen Paaren	$\text{cm}^2 \text{s}^{-1}$
$D^{=}$	Diffusionskoeffizient von zweifach negativ geladenen Paaren	$\text{cm}^2 \text{s}^{-1}$
D_0	präexponentieller Faktor des Diffusionskoeffizienten	$\text{cm}^2 \text{s}^{-1}$
D^0	Diffusionskoeffizient von neutralen Paaren	$\text{cm}^2 \text{s}^{-1}$
d_0	Schichtdicke bei Beginn von Messungen	nm
D_{eff}	effektiver Diffusionskoeffizient	$\text{cm}^2 \text{s}^{-1}$
dG_f^0	freie molare Bildungsenergie bei Normbedingungen (25 °C, 1013 mbar)	kJ mol^{-1}
$d_{interface}$	Dicke des Interface	nm
d_{Oxid}	Oxiddicke	nm
D_P	Diffusionskoeffizient von Phosphor in Silicium	$\text{cm}^2 \text{s}^{-1}$
E_A	Aktivierungsenergie	eV
E_{Cs}	Energie der Cs^+ -Primärionen bei SIMS-Sputtern	keV

Formelzeichen	Beschreibung	Einheit
E_{Ip}	Primärionenenergie	eV
f_{max}	maximaler Anteil eines eingemischten Materials im EMA-Layer	-
f_{min}	minimaler Anteil eines eingemischten Materials im EMA-Layer	-
F_{N_2}	Gasfluss von Stickstoff	slm
F_{O_2}	Gasfluss von Sauerstoff	sccm
G	freie Energie	kJ
H	Enthalpie	kJ
I	elektrischer Strom	A
I_D	Drainstrom am MOSFET	A
I_z	Sekundärionenintensität bei Tiefe z	Pulse/s
J	Fluss (von Dotierstoff)	cm ⁻² s ⁻¹
k	Boltzmannkonstante	1,38062·10 ⁻²³ J K ⁻¹
K	Skalierungsfaktor	-
K_p	Gleichgewichtskonstante	mbar ^r r ist abhängig von Partialdruckverhältnissen und Formelumsatz
L_G	Gate-Länge des MOSFET	nm
ML	Monolage	-
n	Ladungsträgerdichte der Elektronen	cm ⁻³
n	Stoffmenge	mol
N_A	Konzentration der Akzeptoren	cm ⁻³
N_A	Avogadrozahl	6,022·10 ²³ mol ⁻¹
N_D	Konzentration der Donatoren	cm ⁻³
n_e	Elektronen-Ladungsträgerkonzentration	cm ⁻³
n_i	intrinsische Ladungsträgerkonzentration	cm ⁻³
p	Ladungsträgerdichte der Löcher	cm ⁻³
p	Kammerdruck	mbar
P	Gesamtleistung der Lampen	kW
p_{Stoff}	Partialdruck des angegebenen Stoffes	mbar
q	Elementarladung	1,60219·10 ⁻¹⁹ C
q_{Ip}	Primärionenladung	C
R	Widerstand	Ω
R	kinetische Gaskonstante	8,31435 J mol ⁻¹ K ⁻¹
R_S	Flächenwiderstand (Schichtwiderstand)	Ω/sq = Ω
\bar{R}_s	mittlerer Flächenwiderstand (mittlerer Schichtwiderstand)	Ω/sq = Ω
S^0	molare Entropie	J mol ⁻¹ K ⁻¹
t	Zeit	s

Formelzeichen	Beschreibung	Einheit
T	Temperatur	K, °C
T_{10s}	Wafertemperatur nach 10 s Haltezeit	°C
T_{60s}	Wafertemperatur nach 60 s Haltezeit	°C
$t_{Belegung}$	Belegungszeit (Haltezeit bei Belegungstemperatur)	s
t_{Dot}	Dotierzeit (Haltezeit bei Dotiertemperatur)	s
t_{Luft}	Kontaktzeit mit Reinraumatmosfera	min
T_{max}	Maximaltemperatur des Wafers nach Aufheizrampe	°C
$T_{Pin-Thermoelement}$	mittels Pin-Thermoelement gemessene Temperatur	°C
T_{Soll}	Vorgabetemperatur	°C
U	elektrische Spannung	V
U_G	Gate-Spannung am MOSFET	V
U_{Quelle}	Beschleunigungsspannung in der Ionenquelle	V
U_{SD}	Spannung zwischen Source und Drain am MOSFET	V
$U_{Substrat}$	Substratspannung	V
U_{th}	Einsatzspannung des MOSFET	V
V_m	molares Volumen	22,4 l mol ⁻¹ für ideales Gas
V_n	Volumen unter Standardbedingungen	l
w	Fitparameter der Gauß-Funktion (entspricht der Weite des Gauß-Peaks)	cm (nm)
z	Tiefe	nm
z_C	Kontakttiefe von Source und Drain am MOSFET	nm
z_D	Diffusionslänge	nm
z_E	Eindringtiefe von Ionen	nm
z_j	Übergangstiefe	nm
z_o	Fitparameter der Gauß-Funktion (Position des Maximums)	cm (nm)
Δ	ellipsometrischer Winkel	°
ΔH_f	chemische molare Bildungsenthalpie	kJ mol ⁻¹
δ	Einfallswinkel	°
λ	Mischlänge beim Ionenstrahlmischen	nm
μ_n	Elektronenbeweglichkeit	cm ² V ⁻¹ s ⁻¹
μ_p	Löcherbeweglichkeit	cm ² V ⁻¹ s ⁻¹
ρ	spezifischer elektrischer Widerstand	Ω cm
σ	Standardabweichung	%
σ_G	Fitparameter als Maß für Breite eines EMA-Layers mit Gradient	-
σ_{Rs}	Standardabweichung der mittleren Flächenwiderstands	%
φ	Potential	V
Ψ	ellipsometrischer Winkel	°

9. Liste der verwendeten Abkürzungen

Abkürzung	englische Bezeichnung	deutsche Bezeichnung
AES	Auger-electron spectroscopy	Auger-Elektronen-Spektroskopie
ALD	atomic layer doping / atomic layer deposition	Dotieren aus atomaren Schichten / Atomlagen-Abscheidung
APCVD	atmospheric pressure CVD	chemische Gasphasenabscheidung unter Atmosphärendruck
APM	ammonia peroxide mixture	Mischung aus Ammoniumhydroxid und Wasserstoffperoxid
CMOS	complementary metal-oxide-semiconductor	Technologie mit komplementären Metall-Oxid-Halbleitersystemen
CoolMOS	power MOSFET with charge compensation principle	Leistungs-MOSFET Bauform mit Ladungskompensationsprinzip
CVD	chemical vapor deposition	chemische Gasphasenabscheidung
DIBL	drain-induced barrier lowering	Drain-induzierte Barriereerniedrigung
DI-Wasser		de-ionisiertes Wasser
DRAM	dynamic random access memory	dynamischer Halbleiterspeicher mit wahlfreiem Zugang auf den Speicherinhalt
ELA	excimer laser annealing	Heizen mit Excimerlaser
EMA	effective medium approximation	Modell eines effektiven Mediums zur Beschreibung optischer Schichteigenschaften
Fa.		Firma
FD SOI	fully depleted silicon-on-insulator	vollständig verarmtes Silicium auf Isolator
FET	field-effect transistor	Feldeffekttransistor
FinFET	double-gate MOSFET with source/drain regions formed like fins	MOSFET mit doppeltem Gate und flossenförmigen Source-/Drainbereichen
HF	hydrofluoric acid	Fluorwasserstoffsäure, Flusssäure
IC	integrated circuit	integrierte Schaltung
ITRS	International Technology Roadmap for Semiconductors	internationale Planungen der Anforderungen an die Halbleitertechnologie
LD	lethal dose	letale Dosis
LDD	lightly doped drain	gering dotiertes Drain
MAK		maximale Arbeitsplatzkonzentration

Abkürzung	englische Bezeichnung	deutsche Bezeichnung
MOS	metal-oxide-semiconductor	Metall-Oxid-Halbleiter
MOSFET	metal-oxide-semiconductor field-effect transistor	Metall-Oxid-Halbleiter Feldeffekttransistor
MPU	microprocessor unit	Mikroprozessor
MSE	mean square error	mittlerer quadratischer Fehler
NIST-JANAF	National Institute of Standards and Technology- Joint Army Navy Air Force	Tabellen thermochemischer Daten
NMOS	n-channel MOSFET	n-Kanal MOSFET
OED	oxidation-enhanced diffusion	durch Oxidation verstärkte Diffusion
PAI	pre-amorphization implantation	Implantation zur Amorphisierung von Kristallen vor der Dotierstoffimplantation
PGILD	projection-gas-immersion laser doping	gerichtetes Gasimmersions-Laserdotieren
PIII	plasma immersion ion implantation	Plasmainmersions-Ionenimplantation
PinTC	Pin-thermocouple	Thermoelement in Form eine Waferauflagestifts
PMOS	p-channel MOSFET	p-Kanal MOSFET
RCA	Radio Corporation of America	Firmenname
RESURF DMOS	reduced surface field double-diffused MOS	MOS-Struktur mit vermindertem Oberflächenfeld, hergestellt mit doppelter Diffusion
RTA	rapid thermal annealing	Kurzzeit-Ausheizen
RTCVD	rapid thermal chemical vapor deposition	Kurzzeit-Gasphasenabscheidung
RTD	rapid thermal doping	Kurzzeit-Dotieren
RTP	rapid thermal processing	Kurzzeit-Prozessieren
RVD	rapid vapor phase doping	Kurzzeit-Dotieren aus der Gasphase
SC1	Standard Clean 1	Standard-Reinigung 1
SCE	short channel effect	Kurzkanaleffekt
SDE	source/drain extension	Erweiterungsbereich von Source und Drain
SIMS	secondary ion mass spectroscopy	Sekundärionen-Massenspektrometrie
SJ	super junction	Ausführungsform für pn-Übergang
SOI	silicon-on-insulator	Silicium auf Isolator
SPD	solid phase diffusion	Diffusion aus Festkörperphase
TC	thermocouple	Thermoelement
TED	transient enhanced diffusion	vorübergehend verstärkte Diffusion
UHV	ultra-high vacuum	Ultrahochvakuum
ULE	ultra-low energy	(Implantation mit) sehr geringer Energie
ULSI	ultra-large-scale integration	Integrationsgrad: ultrahohe Integrationsdichte von Bauelementen, 100.000-1.000.000 Gate-Äquivalente
USJ	ultra-shallow junction	ultraflacher pn-Übergang
VLSI	very-large-scale integration	Integrationsgrad: sehr hohe Integrationsdichte von Bauelementen, 10.000-100.000 Gate-Äquivalente
VPD	vapor phase doping	Gasphasendotierung

10. Abbildungsverzeichnis

Abbildung 2.1:	Schematische Darstellung eines n-Kanal MOSFET.	11
Abbildung 2.2:	Schematische Schnittdarstellung einer Struktur mit zwei Leiterbahnen zur Definition des Abstandes (Pitch)	17
Abbildung 2.3:	Vergleichende Darstellung wichtiger Kenngrößen der ITRS-Roadmap (2004) über dem Zeitpunkt der Einführung in die Massenfertigung.	18
Abbildung 2.4:	Geforderter Flächenwiderstand über der Übergangstiefe für Source-/Drain-Extensionbereiche planarer MOSFETs mit Jahr der Einführung in die Massenfertigung laut ITRS-Roadmap (2004).	20
Abbildung 3.1:	Temperaturabhängigkeit der Sättigungskonzentration von in Silicium gelöstem Phosphor im Gleichgewicht mit SiP-Präzipitaten (obere Kurve) im Vergleich zur jeweiligen Konzentration der Ladungsträger, die der Konzentration von elektrisch aktivem Phosphor entspricht (untere Kurve) [Solmi '96].	31
Abbildung 3.2:	Vergleich von gaußförmigem Profilverlauf nach Gleichung (18) (dicke Linien) und dem der komplementären Fehlerfunktion nach Gleichung (24) (dünne Linien) für drei verschiedenen Diffusionszeiten für die im Text angegebenen Bedingungen. Dargestellt ist die Phosphorkonzentration über die Tiefe. Für die gaußförmigen Profile sind jeweils die berechneten Flächenwiderstände und die Übergangstiefen bei $C_P = 1 \cdot 10^{18} \text{ cm}^{-3}$ angegeben.	36
Abbildung 3.3:	Temperaturabhängigkeit der intrinsischen Ladungsträgerkonzentration in Silicium nach Gleichung (28) aus [Morin '54].	37
Abbildung 3.4:	Vergleich von Literaturwerten des intrinsischen Diffusionskoeffizienten von Phosphor in Silicium anhand der logarithmischen Darstellung über der reziproken Temperatur (Arrhenius-Darstellung).	39
Abbildung 4.1:	Schematische Darstellung der Anordnung von Komponenten des RTP-Moduls (von oben betrachtet).	40
Abbildung 4.2:	Lage eines Wafers auf drei Auflagepunkten, einer davon mit Thermoelement (Eigenbau), Ansicht von unten durch die Quarzplatte (aus zwei Fotos zusammengesetzt).	41
Abbildung 4.3:	Schematische Darstellung der Anordnung der Lampen und des Quarzgestells der Waferauflage (Ansicht von oben).	42
Abbildung 4.4:	Darstellung einiger relevanter Prozessvariablen (Vorgabe der Temperatur, T_{Soll} , mit Pin-Thermoelement gemessene Temperatur, $T_{\text{Pin-Thermoelement}}$, sowie Gesamtleistung der Lampen, P , Kammerdruck, p_{Kammer} und Gasflüsse von O_2 , F_{O_2} , und N_2 , F_{N_2}) über der Prozesszeit t für den in der Tabelle 4.2 festgelegten Prozess.	45
Abbildung 4.5:	Lage der Auflagepunkte der Wafer in der Prozesskammer (Ansicht von unten). Die Lage des Pin-Thermolements mit jeweiligem Abstand zum Waferrand ist markiert.	50
Abbildung 4.6:	Beispiel des Verlaufs der Solltemperatur eines 900°C -Prozesses.	50

Abbildung 4.7:	Brechungsindex n und Extinktionskoeffizient k von (100)-Silicium in Abhängigkeit von der Wellenlänge nach Jellison.	55
Abbildung 4.8:	Brechungsindex n von Siliciumdioxid in Abhängigkeit von der Wellenlänge als Cauchy-Funktion in dem Bereich der normalen Dispersion (Cauchy-SiO ₂ (thermisch)).	56
Abbildung 4.9:	Schichtmodell für die Ermittlung der Oxidschichtdicken.	56
Abbildung 4.10:	Berechnete Spektren von Δ und Ψ für 0 bis 5 nm Oxiddicke auf Silicium und 70° Einfallswinkel.	57
Abbildung 4.11:	Veränderung von Δ und Ψ in Abhängigkeit von der Oxiddicke für zwei verschiedene Wellenlängen (berechnet).	57
Abbildung 4.12:	Gemessenes Δ -Spektrum von Silicium nach einem HF-Dip zusammen mit entsprechenden besten Approximationen der Dicke von vier verschiedenen dielektrischen Schichten auf reiner Siliciumoberfläche (Ψ -Spektren nicht mit dargestellt, da praktisch identisch).	59
Abbildung 4.13:	Modellierung einer rauen Grenzfläche durch eine angenommene Schicht der Dicke d_R .	60
Abbildung 4.14:	Prinzipaufbau der Vierspitzen-Messanordnung.	63
Abbildung 5.1:	Gemessene Spektren von Ψ und Δ für den gesamten zugänglichen Messbereich von 250-870 nm unmittelbar nach HF-Dip eines mit Caro'scher Säure gereinigten Siliciumwafers im Vergleich mit zu erwartenden Verläufen, ermittelt aus bestem Fit für den Bereich von 340-800 nm.	74
Abbildung 5.2:	Spektren von Ψ und Δ für Wafer A-1 unmittelbar nach einem HF-Dip und berechnete Werte für besten Fit mit 0,57 nm Oxiddicke.	75
Abbildung 5.3:	Anstieg der Schichtdicke von Wafer A-3 an Luft nach einer HF-Behandlung, Darstellung im Logarithmus der Zeit an Luft. Die kleine Darstellung zeigt den Verlauf mit linearer Zeitachse).	78
Abbildung 5.4:	Ellipsometrische Messwerte im Bereich von 400-600 nm an Si-Waferoberflächen nach Caro-Reinigung (links) und unmittelbar nach HF-Behandlung (rechts) zusammen mit den theoretischen Verläufen aus der besten Anpassung des Modells mit einer Oxidschicht (rot) und mit zusätzlichem Interface aus amorphem Silicium (blau). Rechts mit dargestellt ist Verlauf von Δ bei idealer, reiner Si-Oberfläche.	79
Abbildung 5.5:	Vergleich der besten Fits bei Verwendung des Modells mit einer amorphen Zwischenschicht, bestimmt unter Berücksichtigung von unterschiedlichen Wellenlängenbereichen	82
Abbildung 5.6:	Zusammenhang von vorgegebenem Temperatursollwert, dem mit PinTC gemessenen Istwert und der Leistungsaufnahme aller Lampen während eines typischen Prozesses unter Einsatz eines neuen PinTCs.	83
Abbildung 5.7:	Temperaturverläufe und zugehörige Lampenleistungen vor dem Ausfall eines PinTC.	84
Abbildung 5.8:	Vergleich der Gesamtleistungsaufnahme bei verschiedenen Prozessen mit 900°C Soll-Haltetemperatur unter Einsatz verschiedener Pin-Thermoelemente.	85
Abbildung 5.9:	Vergleich der Leistungsaufnahme bei verschiedenen Heizrampen auf 900 °C Solltemperatur bei Verwendung eines Pin-Thermoelements ohne Verkapselung.	86
Abbildung 5.10:	Vergleich der Reproduzierbarkeit der benötigten Lampenleistung bei Verwendung des Pin-Thermoelements mit Verkapselung.	86
Abbildung 5.11:	Darstellung der Temperaturverläufe, gemessen in der Wafermitte mittels TC-Wafer (obere Kurven), und der zugehörigen Verläufe der Gesamtleistung der Lampen (untere Kurven) für verschiedene Testprozesse im Vergleich zu Leistungsverläufen dreier Dotierprozesse. Die jeweiligen Solltemperaturen, die den mittels Pin-Thermoelement gemessenen Temperaturen entsprechen, sind angegeben.	87

- Abbildung 5.12: Darstellung der Temperaturverläufe, gemessen in der Wafermitte mittels TC-Wafer, bei jeweils angegebener konstanter Solltemperatur (für 60 s Haltezeit). Für den Testprozess bei 920 °C Solltemperatur ist die Bestimmung der Maximaltemperatur (T_{max}), der Temperatur T_{10s} (entsprechend dem Ende eines 10 s Dotierprozesses) und der Temperatur T_{60s} angegeben. 88
- Abbildung 5.13: Darstellung eines Beispiels der Temperaturermittlung über die aufgewendete Lampenleistung. Angegeben sind die mittels TC-Wafer gemessenen Temperaturen und die bei diesen Testprozessen benötigten Lampenleistungen für verschiedene Prozesszeiten (nach 10 s und nach 60 s) in Abhängigkeit von der eingestellten Solltemperatur (Temperatur des Pin-Thermoelements). Zum Vergleich sind die bei drei Dotierprozessen benötigten Lampenleistungen angegeben (nach 10 s Haltezeit) und die daraus zu erwarteten tatsächlichen Temperaturen (schwarze Pfeile) sowie die Leistung eines Oxidationsprozesses bei einer Solltemperatur von 950°C unter Verwendung eines Original-PinTCs (orange Pfeile). 89
- Abbildung 5.14: Oxidschichtdickendarstellung im Höhengitternetz eines Wafers nach statischer Oxidation ohne Leistungskorrektur. 90
- Abbildung 5.15: Schichtdickenverlauf bei statischer Oxidation des Wafers aus Abbildung 5.14 für die links unten dargestellte Messlinie. 91
- Abbildung 5.16: Schichtdickenverlauf bei stationärer Oxidation bei 5 mbar mit gleicher Leistung wie beim Wafer aus Abbildung 5.15 für die links unten dargestellte Messlinie. 91
- Abbildung 5.17: Darstellung der Korrekturfaktoren der Lampen der bezeichneten Position für eine verbesserte Schichtdickenverteilung. 92
- Abbildung 5.18: Schichtdickenverlauf bei statischer Oxidation mit Leistungskorrektur für die links unten dargestellte Messlinie. 93
- Abbildung 5.19: Schichtdickenverlauf bei stationärer Oxidation mit Leistungskorrektur für die links unten dargestellte Messlinie. 93
- Abbildung 5.20: Oxidschichtdicke in Abhängigkeit von der Wafertemperatur für statische Oxidation bei 40 mbar ohne Leistungskorrektur. 94
- Abbildung 5.21: Temperaturverlauf bei statischer Oxidation mit Leistungskorrektur für die links unten dargestellte Messlinie, ermittelt aus dem Schichtdickenverlauf aus Abbildung 5.18 und dem in Abbildung 5.20 dargestellten Zusammenhang zwischen Schichtdicke und Wafertemperatur. 95
- Abbildung 5.22: Darstellung der aus den Oxidschichtdickenverteilungen ermittelten lokalen Wafertemperaturen während eines statischen Oxidationsprozesses unter 40 mbar Sauerstoff (Ansicht der Waferoberseite). Die Farben entsprechen den in der Legende angegebenen Temperaturen. Eingezeichnet sind weiterhin Linien von Isothermen, wobei die grüne Linie den Temperaturmittelwert darstellt und der Abstand der dünnen Linien ein Temperaturintervall von 0,1 % angibt. Die Lage der Waferauflagepunkte ist durch rote Kreise markiert. 96
- Abbildung 5.23: Wafertemperaturdarstellung im Höhengitternetz (je höher der Netzpunkt, je höher die Temperatur) des Wafers aus Abbildung 5.22. Ansicht der Waferoberseite. Eingezeichnet ist der deutliche Einfluss der Auflagepunkte auf die Temperatur. 96
- Abbildung 5.24: Bezug der Position der Lampen zu den Temperaturschwankungen. Die Stablampen sind symbolisch im Schnitt dargestellt. 97
- Abbildung 5.25: Mittlerer Flächenwiderstand von vier fortlaufend unter gleichen Bedingungen prozessierten Wafern ($T_{Soll} = 850$ °C, Rampe 60/30/60 s und $p = 9$ mbar). 100
- Abbildung 5.26: Darstellung der Gleichgewichtskonstanten K_P der Zerfallsreaktion von PH_3 (g) in Abhängigkeit von der reziproken Temperatur. Berechnet aus Daten von [Lodders '99] für 1000 mbar, oberhalb von 1180 K wurde dort der Phosphorreferenzzustand P_2 (g) wie bei der Betrachtung von Gleichung (62) benutzt. 102

- Abbildung 5.27: Aus der Gleichgewichtsbetrachtung der homogenen Gasphasenreaktion (62) ermittelter Anteil von zerfallenem Phosphin bei einer Temperatur von 1000 °C in Abhängigkeit vom eingesetzten Phosphinpartialdruck in Wasserstoffatmosphäre bei drei verschiedenen Gesamtdrücken. Angegeben ist der bei Verwendung von 50 vpm PH₃ in H₂ bei dem jeweiligen Gesamtdruck vorherrschende anfängliche Phosphinpartialdruck. 103
- Abbildung 5.28: Abhängigkeit der Oberflächenkonzentration von Phosphor in Silicium nach [Sato '98] für eine Gasphasendiffusion von Phosphin in Wasserstoff (1000 °C, 3 min) extrapoliert auf die Partialdruckbereiche der durchgeführten Experimente mit 50 vpm PH₃ in H₂. Zum Vergleich sind Ergebnisse von [Kiyota '94] für eine Gasphasendiffusion unter Atmosphärendruck bei 900°C und 30 min mit angegeben. 105
- Abbildung 5.29: Typischer Ablauf des Zwei-Stufen-Dotierverfahrens der Prozessvariante 1 mit einem Ausheizschritt vor der Belegung, dargestellt anhand des Verlaufs der Solltemperatur. 108
- Abbildung 5.30: Typischer Ablauf des Zwei-Stufen-Dotierverfahrens der Prozessvariante 2 ohne Ausheizschritt vor der Belegung, dargestellt anhand des Verlaufs der Solltemperatur. 108
- Abbildung 5.31: Darstellung des Verlaufs des Flächenwiderstands über den Waferdurchmesser (rechts unten symbolisiert), Eintreibprozess bei 1000 °C Maximaltemperatur (ermittelt aus Lampenleistung) und 60 s Dotierzeit (Wafer 3 in Abbildung 5.33). Die Lage des Verlaufs bezüglich der Lampen ist wie bei den Oxidationsversuchen. Die Angaben zur Statistik beziehen sich auf den gesamten Testdurchmesser von ca. 130 mm. 111
- Abbildung 5.32: Darstellung des Verlaufs des Flächenwiderstands über den Waferdurchmesser (rechts unten symbolisiert), Eintreibprozess bei 1032 °C Maximaltemperatur (ermittelt aus Lampenleistung) und 60 s Dotierzeit (Wafer 9 in Abbildung 5.33). Der Flat war hier um 90° gegenüber Wafer 3 gedreht, jedoch ist der gleiche Widerstandsverlauf bezüglich der Lampenlage dargestellt. Die Angaben zur Statistik beziehen sich auf den gesamten Testdurchmesser von ca. 130 mm. 111
- Abbildung 5.33: Reproduzierbarkeit der Dotierergebnisse für 14 gleichartig durchgeführte Dotierprozesse. Dargestellt sind die Mittelwerte der gemessenen Flächenwiderstände in einem homogener dotierten Bereich von 75 mm Durchmesser in der Mitte der Wafer und die für diese Versuche aus dem aufgewendeten Lampenleistungen ermittelten Maximaltemperaturen. (Die Versuche wurden an verschiedenen Tagen durchgeführt, die Verbindungslinien symbolisieren keinen Zusammenhang, sondern sollen nur den Vergleich zu den Prozesstemperaturen erleichtern.) 112
- Abbildung 5.34: Darstellung des Verläufe des Flächenwiderstands über dem Waferdurchmesser (rechts unten symbolisiert) für die links bezeichneten Wafer mit Wafernummern entsprechend Tabelle 5.4. Die Lage der Verläufe ist wie bei Oxidationsversuchen. 113
- Abbildung 5.35: Flächenwiderstände in Abhängigkeit vom Ort auf der Waferoberfläche in einer räumlichen Gitterdarstellung, die Höhe entspricht einem Maß für den jeweiligen Flächenwiderstand, und jeweils dazugehörige „Höhenliniendarstellung“ mit Linien gleichen Flächenwiderstands für drei verschiedene Wafer (siehe Tabelle 5.4). 116
- Abbildung 5.36: Flächenwiderstand (Mittelwert, siehe Tabelle 5.4) in Abhängigkeit von der Maximaltemperatur des Eintreibschritts für gleiche Eintreibzeit von 10 s. Die Gerade wurde aus linearer Regression ermittelt und ermöglicht die Abschätzung der Streuung der Prozessergebnisse. 118
- Abbildung 5.37: Flächenwiderstand in Abhängigkeit von der Eintreibzeit bei gleicher Temperatur (mit angegeben) in 5 mbar Sauerstoff. Die Angabe der Verbindungslinie erfolgte hier zur günstigeren Darstellung des Zusammenhangs, eine funktionelle Abhängigkeit kann aus den Ergebnissen noch nicht abgeleitet werden. 119
- Abbildung 5.38: Flächenwiderstand in Abhängigkeit von der Zeit der Phosphineinwirkung bei gleicher Belegungstemperatur und gleichen Eintreibbedingungen. Für 10 s Belegungszeit sind zwei Datenpunkte mit demselben Widerstandswert eingezeichnet (Waferbezeichnungen entsprechend Tabelle 5.5). 120
- Abbildung 5.39: Einfluss der maximalen Eintreibtemperatur auf den Flächenwiderstand der unterschiedlich belegten Wafer (Waferbezeichnungen entsprechend Tabelle 5.5). 121

Abbildung 5.40:	Flächenwiderstand gemessen in Wafermitte in Abhängigkeit von der Zeit des Waferkontakts mit Reinraumluft nach dem Dotierprozess (Wafer aus Prozessvariante 1).	122
Abbildung 5.41	Anstieg des Flächenwiderstands von Wafer 3 nach der Entnahme aus der Vakuumschleuse nach dem Dotierprozess bei einer Messung mit aufgesetzten Messspitzen an Luft. In dem eingefügten Diagramm ist die Zeitskala zum Vergleich logarithmisch dargestellt.	123
Abbildung 5.42:	Vergleich des Anstiegs des Flächenwiderstands eines mit Prozessvariante 2 dotierten Wafers mit dem Anstieg einer mittels Ellipsometer gemessenen Schichtdicke desselben Wafers jeweils in Luft nach einer Oxidentfernung durch HF-Behandlung.	124
Abbildung 5.43:	Flächenwiderstand in Abhängigkeit von der Dicke des wachsenden Oxids.	125
Abbildung 5.44:	Differenz zwischen dem gemessenen Flächenwiderstand und dem linearen Fit aus Abbildung 5.43 über der erwarteten Oxidschichtdicke (oben) bzw. der Tiefe des Siliciumumsatzes durch die Oxidation. Die Gitterlinien kennzeichnen die Periodizität der Oszillationen, die dazu angegebenen Werte beziehen sich dabei auf die Tiefe des Siliciumumsatzes. Der eingezeichnete Verlauf einer gedämpften Sinusfunktion soll hier nur als Referenz dienen.	126
Abbildung 5.45:	Differenz von der gemessenen Oxidschichtdicke und der nach dem logarithmischen Gesetz (71) zu erwartenden Schichtdicke (Oxidwachstum und Fit aus Abbildung 5.42) in Abhängigkeit von der erwarteten Schichtdicke. Einige offensichtlich stark gestreuten Messwerte wurden hier für eine bessere Anschaulichkeit nicht mit dargestellt. In der oberen Skala ist zum Vergleich die Oxidationszeit an Luft angegeben. Die Gitterlinien zeigen die Periodizität der Schwankungen auf der Oxiddickenskala. Der Verlauf der gedämpften Sinusfunktion ist nur als Referenz angegeben.	127
Abbildung 5.46:	Vergleich von mit SIMS gemessenen Profilen von Phosphor in Silicium für zwei verschiedene Primärionenenergien und zwei verschiedene Dotierungen, links ist die ermittelte Phosphorkonzentration jeweils linear aufgetragen.	128
Abbildung 5.47:	Vergleich von SIMS-Profilen einer Probe, die an verschiedenen Anlagen unter Verwendung von verschiedenen Primärionenenergien aufgenommen wurden.	129
Abbildung 5.48:	Vergleich der aus den SIMS-Profilen ermittelten Dosis in Abhängigkeit von der für die Messung verwendeten Primärionenenergie.	130
Abbildung 5.49:	Vergleich der Dotierprofile (Konzentration von P in Si) von unterschiedlichen Bereichen eines Wafers, dotiert mittels Prozessvariante 1. Angegeben sind die Flächenwiderstände und die Flächenkonzentrationen von Phosphor von den entsprechenden Stellen.	131
Abbildung 5.50:	Vergleich von Phosphorprofilen in Silicium von mit den zwei unterschiedlichen Prozessvarianten dotierten Wafern.	131
Abbildung 5.51:	Einfluss von Variationen von Prozessbedingungen der Dotiervariante 2 auf die Verteilung des Phosphors im Silicium.	132
Abbildung 5.52:	Mittels SIMS gemessene Konzentrationen von Phosphor in Silicium nach der Zwei-Stufen-Dotierung bei verschiedenen Eintreibetemperaturen. Die entsprechenden angenäherten Gauß-Verläufe sind gestrichelt dargestellt.	133
Abbildung 5.53:	Abhängigkeit des gemessenen Flächenwiderstands von der maximalen Eintreibetemperatur im Vergleich zu den aus den SIMS-Profilen ermittelten Werten. (Schwarz dargestellt $t_{Dot} = 10$ s, rot $t_{Dot} = 40$ s und HF-Dip vor SIMS-Messung, die dünne Linie zeigt den Zusammenhang aus den Mittelwerten des Flächenwiderstands für 120 mm Testdurchmesser aus Abbildung 5.36).	135
Abbildung 5.54:	Zusammenfassende Darstellung der aus den SIMS-Profilen ermittelten Konzentrationen (Oberflächenkonzentration C_{OF} des aktiven Phosphors aus Gauß-Verlauf und Flächenkonzentration C_D [Dosis] aus Integration der gemessenen Profile und der Gauß-Verläufe). Die farbig umrandeten Punkte sind Ergebnisse von SIMS-Analysen von HF behandelten Proben (d.h. ohne Pile-up, siehe unten), die Farben der eingezeichneten Linien der ermittelten Zusammenhänge beziehen sich auf die Füllfarben der entsprechenden Datenpunkte.	136

Abbildung 5.55:	Zusammenhang zwischen Flächenwiderstand und Übergangstiefe von im zweistufigen Prozess dotierten Wafern im Vergleich zu den geforderten Werten laut Roadmap (siehe Abbildung 2.4) und errechneten Werten unter Annahme einer idealen Eindiffusion von einer Monolage Phosphor (siehe Abbildung 3.2).	137
Abbildung 5.56:	Vergleich der aus den Profilen ermittelten Diffusionskoeffizienten mit den intrinsischen aus Abbildung 3.4.	138
Abbildung 5.57:	SIMS-Tiefenprofile von ^{31}P , ^{16}O und ^{29}Si nahe der Oberfläche eines Siliciumwafers nach dem Zwei-Stufen-Dotierverfahren bei $T_{max} = 1165\text{ °C}$ und $t_{Dot} = 10\text{ s}$. Die Gauß-Näherung des Phosphorprofils ist gestrichelt dargestellt.	139
Abbildung 5.58:	SIMS-Tiefenprofile von Phosphor in Silicium nach der Zwei-Stufen-Dotierung und unterschiedlicher HF-Behandlung vor den Analysen (SIMS 1: $E_{Cs} = 3,5\text{ keV}$, SIMS 2: $E_{Cs} = 0,5\text{ keV}$). Angegeben sind die Anzahl der HF-Behandlungen und die mittels Ellipsometer bestimmten Schichtdicken. Der Bereich nahe der Oberfläche mit erhöhter Phosphorkonzentration ist oben rechts vergrößert dargestellt.	141
Abbildung 5.59:	Vergleich des Phosphorprofils in Silicium eines mittels Prozessvariante 2 dotierten Wafers mit denen von Wafern, womit nur eine Belegung bzw. eine Belegung mit vorherigem Ausheizen in Dotiergasatmosphäre durchgeführt wurde. Das eingefügte Diagramm zeigt vergrößert den Bereich bis zu einer Tiefe von 20 nm.	142
Abbildung 5.60:	AES-Spektren zweier Wafer, bei denen nur die Belegung aber kein Eintreiben erfolgte. Der Unterschied ist eine HF-Behandlung des einen Wafers vor der Messung, das eingefügte Diagramm zeigt einen Teil des Rohspektrums des nicht mit HF behandelten Wafers vor der Ableitung.	143
Abbildung 5.61:	Vergleich der AES-Spektren aus Abbildung 5.60 mit eingezeichneten Peaklagen.	143
Abbildung 5.62:	Beispiel für ein ellipsometrisches Δ -Spektrum von Wafer 3 (ca. Wafermitte) im Vergleich zum Referenzspektrum und Vergleich der jeweils günstigsten Approximation dreier Modelle.	144
Abbildung 5.63:	Verwendete Modelle für die Auswertung der ellipsometrischen Spektren.	145
Abbildung 5.64:	Darstellung der ermittelten Gradienten der jeweiligen Interface-Modelle für den Fit aus Abbildung 5.62.	145
Abbildung 5.65:	Delta-Spektrum von Wafer 1 gemessen an verschiedenen Stellen im Vergleich zum Spektrum eines HF-gereinigten Wafers.	147
Abbildung 5.66:	Vergleich ausgewählter Δ -Spektren von Wafer 1 mit den berechneten von günstigsten Modellen.	147
Abbildung 5.67:	Vergleich ausgewählter Δ -Spektren verschiedener dotierter Wafer (gemessen in der Wafermitte) mit dem Referenzspektrum. Blau dargestellt: berechnete Spektren aus günstigsten Modellen.	148
Abbildung 5.68:	Vergleich der Δ -Spektren von dotierten Wafer (gemessen in der Wafermitte) mit und ohne HF-Behandlung nach dem Dotierprozess mit dem Referenzspektrum. Blau dargestellt: berechnete Spektren aus günstigsten Modellen.	150
Abbildung 5.69:	Verlauf der gemessenen Oxiddicke (offene Punkte) über dem Waferdurchmesser dreier bei unterschiedlicher Eintreibetemperatur dotierter Wafer im Vergleich zu den Flächenwiderständen.	151

11. Tabellenverzeichnis

Tabelle 3.1:	Parameter des intrinsischen Diffusionskoeffizienten von Phosphor in Silicium.	38
Tabelle 3.2:	Parameter der extrinsischen Diffusionskoeffizienten von Phosphor in Silicium.	39
Tabelle 4.1:	Verwendete Gaslinien und der jeweils maximal mittels Massendurchflussmesser zu regelnde Gasfluss.	43
Tabelle 4.2:	Rezept in tabellarischer Form für ein 10 s - Annealing in 5 mbar Sauerstoff bei 900 °C Solltemperatur.	44
Tabelle 4.3:	Übersicht über die protokollierten Prozessparameter.	44
Tabelle 4.4:	Ablauf der Untersuchungen zur nasschemischen Reinigung und Terminierung der Waferoberflächen und zum Wachstumsverhalten natürlicher Oxidschichten.	47
Tabelle 4.5:	Übersicht über die Reinigungs- und Oxidentfernungsprozeduren.	48
Tabelle 4.6:	Änderung von Δ und Ψ durch eine 0,1 nm dicke Oxidschicht gegenüber unbedeckter Siliciumoberfläche für zwei verschiedene Wellenlängen bei einem Einfallswinkel von 70°.	58
Tabelle 4.7:	Ermittelte Schichtdicken und -fehler aus Auswertung des Spektrums von Abbildung 4.12 unter Annahme der angegebenen Schichten auf reinem Silicium (Cauchy-Schichtmodelle aus SpectraRay).	58
Tabelle 4.8:	Eigenschaften der vorhandenen Vierspitzen-Messköpfe.	66
Tabelle 5.1:	Zusammenfassung der ermittelten (äquivalenten) Oxiddicken nach verschiedenen Vorbehandlungsschritten an Siliciumwafern	75
Tabelle 5.2:	Vergleich dreier Schichtmodelle für die Auswertung ellipsometrischer Messungen im Wellenlängenbereich von 400-600 nm nach verschiedenen Oberflächenbehandlungen von Silicium (BG-Interf. = Brüggemann-Interface).	80
Tabelle 5.3:	Ermittelte Wafertemperaturen und Aufheizrampen von Dotierprozessen.	89
Tabelle 5.4:	Ausgewählte Dotierexperimente der Untersuchungen mit Prozessvariante 2 des Zweistufenprozesses. Die Lampenheizung erfolgte mit Leistungskorrektur, vor dem Eintreiben erfolgte eine Vorheizphase von $T_{\text{Soll}} = 550$ °C für 60 s, die Gasdrücke wurden durch stationären Gassfluss eingestellt, außer beim Eintreiben von Wafer 1. Die Ermittlung der maximalen Prozesstemperatur erfolgte aus der Auswertung der aufgewendeten Lampenleistungen, die Angaben zum Flächenwiderstand beziehen sich auf einen Testdurchmesser von 120 mm (47 Messpunkte). Die Erklärung der Abkürzungen von Prozessvariationen ist unten angegeben.	114
Tabelle 5.5:	Übersicht über Dotierversuche mit verschiedenen Belegungsschritten und gleichen Eintreibbedingungen nach Prozessvariante 2, Angabe von Haltezeit während der Belegungstemperatur, maximaler Temperatur während des Eintreibens (ermittelt aus aufgewendeten Lampenleistungen) und mittlerem Flächenwiderstand (120 mm Testdurchmesser).	120

Tabelle 5.6:	Zusammenfassung ausgewählter Ergebnisse von Dotierexperimenten mit dem Zwei-Stufen-Prozess. Die Ergebnisse wurden aus den gemessenen Phosphor-Konzentrationsverläufen (SIMS) und aus den dem jeweiligen Gauß-Fit entsprechenden Konzentrationen (Gauß) ermittelt. Einige Wafer wurden vor der SIMS-Analyse mit verdünnter Flusssäure (HF, 1%, 3 min) behandelt.	134
Tabelle 5.7:	Übersicht über die aus den Approximationen ermittelten Modellparameter für die in Abbildung 5.62 dargestellten Spektren.	146
Tabelle 5.8:	Übersicht über die aus den Approximationen ermittelten Modellparameter für die in Abbildung 5.66 dargestellten Spektren.	148
Tabelle 5.9:	Übersicht über die aus den Approximationen der ellipsometrischen Spektren ermittelten Modellparameter für verschiedene dotierte Wafer (gleiche Wafernummern wie in Tabelle 5.4).	149

12. Literatur

- [Agarwal '00] Agarwal, A. (2000). *Ultra-shallow junction formation using conventional ion implantation and rapid thermal annealing*. Conference on Ion Implantation Technology, Alpbach, Austria, Ed. IEEE, 293-299.
- [Angermann '04] Angermann, H., W. Henrion, M. Rebien und A. Roseler (2004). *Wet-chemical preparation and spectroscopic characterization of Si interfaces*. Appl. Surf. Sci. **235** (3), 322-339.
- [Archer '65] Archer, R. J. und G. W. Gobeli (1965). *Measurement of Oxygen Adsorption on Silicon by Ellipsometry*. J. Phys. Chem. Solids **26**, 343-351.
- [Behrendt '93] Behrendt, W., D. Koschel, L. Gmelin, R. J. Meyer, E. H. E. Pietsch und E. Fluck, (1993). *Gmelin handbook of inorganic and organometallic chemistry, Phosphorus, Supplement Volume C1, Mononuclear compounds with hydrogen*. Berlin u.a., Verl. Chemie u.a.
- [Bozso '91] Bozso, F. und P. Avouris (1991). *Adsorption of phosphorus on Si(111): Structure and chemical reactivity*. Phys. Rev. B: Condens. Matter **43** (2), 1847-1850.
- [Chase '98] Chase, M. W. (1998). *NIST-JANAF thermochemical tables*. Woodbury, NY, American Institute of Physics.
- [Cho '92] Cho, B.-J., S.-K. Park und C.-K. Kim (1992). *Estimation of effective diffusion time in a rapid thermal diffusion using a solid diffusion source*. IEEE Trans. Electron Devices **39** (1), 111-117.
- [Choi '03] Choi, H., T. Tanabe, N. Kotaki, K. W. Koh, J. C. Shim, K. T. Park, H. Kurino und M. Koyanagi (2003). *New Silicon-on-Insulator (SOI) Flash Memory with Side Channel and Side Floating Gate*. Jpn. J. Appl. Phys., Part 1 **42** (6A), 3361-3363.
- [Chow '83] Chow, R. und Y. G. Chai (1983). *A PH₃ cracking furnace for molecular beam epitaxy*. J. Vac. Sci. Technol., A **1** (1), 49-54.
- [Christensen '03] Christensen, J. S., H. H. Radamson, A. Y. Kuznetsov und B. G. Svensson (2003). *Phosphorus and boron diffusion in silicon under equilibrium conditions*. Appl. Phys. Lett. **82** (14), 2254-2256.
- [Clarysse '02] Clarysse, T., D. Vanhaeren und W. Vandervorst (2002). *Impact of probe penetration on the electrical characterization of sub-50 nm profiles*. J. Vac. Sci. Technol., B **20** (1), 459-466.
- [Claverie '99] Claverie, A., L. F. Giles, M. Omri, B. De Mauduit, G. Ben Assayag und D. Mathiot (1999). *Nucleation, growth and dissolution of extended defects in implanted Si: Impact on dopant diffusion*. Nucl. Instrum. Methods Phys. Res., Sect. B **147** (1-4), 1-12.

- [de Chambost '04] de Chambost, E., A. Merkulov, M. Schuhmacher und P. Peres (2004). *Accurate in depth profiling of As and P shallow implants by secondary ion mass spectroscopy*. J. Vac. Sci. Technol., B **22** (1), 341-345.
- [Den '06] Den, W., H. Bai und Y. Kang (2006). *Organic airborne molecular contamination in semiconductor fabrication clean rooms*. J. Electrochem. Soc. **153** (2), G149-G159.
- [Downey '98] Downey, D. F., J. W. Chow, W. Lerch, J. Niess und S. D. Marcus (1998). *Effects of small concentrations of oxygen in RTP annealing of low energy boron, BF₂ and arsenic ion implants*. Mater. Res. Soc. Symp. Proc., **525**, 263-271.
- [Eto '05] Eto, T. und K. Shibahara (2005). *Accuracy of secondary ion mass spectrometry depth profiling for sub-keV As⁺ implantation*. Jpn. J. Appl. Phys., Part 1 **44** (4 B), 2433-2436.
- [Fahey '89] Fahey, P. M., P. B. Griffin und J. D. Plummer (1989). *Point defects and dopant diffusion in silicon*. Reviews of Modern Physics **61** (2), 289-384.
- [Faifer '05] Faifer, V. und M. Current (2005). *Ultrashallow Junctions Demand Better Doping Metrology*. Semicond. Int. **28** (12), 38-44.
- [Fehlner '70] Fehlner, F. P. und N. F. Mott (1970). *Low-temperature oxidation*. Oxid. Met. **2** (1), 59-99.
- [Foggiato '05] Foggiato, J. und W. S. Yoo (2005). *Millisecond flash annealing: Applications for USJ formation and optimization of device electrical characteristics*. Mater. Sci. Eng., B **124-125** (SUPPL.), 219-222.
- [Ghoshtagore '71] Ghoshtagore, R. N. (1971). *Intrinsic Diffusion of Boron and Phosphorus in Silicon Free From Surface Effects*. Phys. Rev. B: Condens. Matter **3** (2), 389-396.
- [Glenn '00] Glenn, J. und J. Siekkinen (2000). *A novel vertical deep trench RESURF DMOS (VTR-DMOS)*. IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD), 197-200.
- [Gopalan '03] Gopalan, C., P. S. Chakraborty, J. Yang, T. Kim, Z. Wu, M. R. McCartney, S. M. Goodnick, M. N. Kozicki und T. J. Thornton (2003). *Shallow source/drain extensions for deep submicron MOSFETs using spin-on-dopants*. IEEE Trans. Electron Devices **50** (5), 1277-1283.
- [Gossmann '03] Gossmann, H.-J. L. (2003). *Lateral Abruptness of the Source/Drain Extension, Doping Supplement of the ITRS*, <http://public.itrs.net>.
- [Gräf '89] Gräf, D., M. Grundner und R. Schulz (1989). *Reaction of water with hydrofluoric acid treated silicon(111) and (100) surfaces*. J. Vac. Sci. Technol., A **7** (3), 808-813.
- [Griffin '95] Griffin, P. B., S. W. Crowder und J. M. Knight (1995). *Dose loss in phosphorus implants due to transient diffusion and interface segregation*. Appl. Phys. Lett. **67** (4), 482-484.
- [Habuka '03] Habuka, H., S. Ishiwari, H. Kato, M. Shimada und K. Okuyama (2003). *Airborne Organic Contamination Behavior on Silicon Wafer Surface*. J. Electrochem. Soc. **150** (2), G148-G154.
- [Haddara '00] Haddara, Y. M., B. T. Folmer, M. E. Law und T. Buyuklimanli (2000). *Accurate measurements of the intrinsic diffusivities of boron and phosphorus in silicon*. Appl. Phys. Lett. **77** (13), 1976-1978.
- [Hattori '01] Hattori, Y., T. Suzuki, M. Kodama, E. Hayashii und T. Uesugi (2001). *Shallow angle implantation for extended trench gate power MOSFETs with super junction structure*. IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD), 427-430.

- [Herzinger '98] Herzinger, C. M., B. Johs, W. A. McGahan, J. A. Woollam und W. Paulson (1998). *Ellipsometric determination of optical constants for silicon and thermally grown silicon dioxide via a multi-sample, multi-wavelength, multi-angle investigation*. J. Appl. Phys. **83** (6), 3323-3336.
- [Holleman '95] Holleman, A. F. und N. Wiberg (1995). *Lehrbuch der anorganischen Chemie*. Berlin [u.a.], De Gruyter.
- [Hsueh '70] Hsueh, Y. W. (1970). *Phosphorus Diffusion in Silicon Using Phosphine*. J. Electrochem. Soc. **117** (6), 807-811.
- [Hu '97] Hu, Y. Z., S. P. Tay, Y. Wasserman, C. Y. Zhao, K. J. Hebert und E. A. Irene (1997). *Characterization of ultrathin SiO₂ films grown by rapid thermal oxidation*. J. Vac. Sci. Technol., A **15** (3), 1394-1398.
- [Huff '04] Huff, H. R. und P. M. Zeitzoff (2004). *An analytical look at vertical transistor structures*. Solid State Tech. **47** (8), 59-66+72.
- [Hutchby '02] Hutchby, J. A., G. I. Bourianoff, V. V. Zhirnov und J. E. Brewer (2002). *Extending the Road Beyond CMOS*. IEEE Circuits & Devices Magazine **18** (3), 28-41.
- [ITRS '06] ITRS (2006). *International Technology Roadmap for Semiconductors*, <http://public.itrs.net>.
- [Iwai '99] Iwai, H. (1999). *CMOS Technology-Year 2010 and Beyond*. IEEE Journal of Solid-State Circuits **34** (3), 357-366.
- [Jones '98] Jones, E. C. und E. Ishida (1998). *Shallow junction doping technologies for ULSI*. Mater. Sci. Eng., R **24** (1-2), 1-80.
- [Kamiura '93] Kamiura, Y., Y. Mizokawa, M. Iida, Y. Isobe und K. Kawamoto (1993). *Investigation of Reoxidation and Phosphorus Behavior in HF-Treated Heavily Doped Silicon Surfaces*. Jpn. J. Appl. Phys., Part 1 **32** (11A), 4863-4869.
- [Kaneta '99] Kaneta, C., T. Yamasaki, T. Uchiyama, T. Uda und K. Terakura (1999). *Structure and electronic property of Si(100)/SiO₂ interface*. Microelectron. Eng. **48** (1-4), 117-120.
- [Kasi '91] Kasi, S. R., M. Liehr, P. A. Thiry, H. Dallaporta und M. Offenbergl (1991). *Hydrocarbon reaction with HF-cleaned Si(100) and effects on metal-oxide-semiconductor device quality*. Appl. Phys. Lett. **59** (1), 108-110.
- [Kim '90] Kim, J.-G., B.-J. Cho und C.-K. Kim (1990). *AES study of rapid thermal boron diffusion into silicon from a solid diffusion source in oxygen ambient*. J. Electrochem. Soc. **137** (9), 2857-2860.
- [Kim '00] Kim, K.-S., Y.-H. Song, K.-T. Park, H. Kurino, T. Matsuura, K. Hane und M. Koyanagi (2000). *A novel doping technology for ultra-shallow junction fabrication: boron diffusion from boron-adsorbed layer by rapid thermal annealing*. Thin Solid Films **369** (1-2), 207-212.
- [Kim '87] Kim, K.-T. und C.-K. Kim (1987). *Formation of Shallow p⁺-n Junctions Using Boron-Nitride Solid Diffusion Source*. Electron device letters **12**, 569-571.
- [Kim '02] Kim, S.-D., C.-M. Park und J. C. S. Woo (2002). *Advanced source/drain engineering for box-shaped ultrashallow junction formation using laser annealing and pre-amorphization implantation in sub-100-nm SOI CMOS*. IEEE Trans. Electron Devices **49** (10), 1748-1754.
- [Kim '95] Kim, S. Y. und E. A. Irene (1995). *An evaluation of errors in determining the refractive index and thickness of thin SiO₂ films using a rotating analyzer ellipsometer*. Rev. Sci. Instrum. **66** (11), 5277-5281.

- [King '05] King, T.-J. (2005). *Taking silicon to the limit: Challenges and opportunities*. Electrochem. Soc. Interface **14** (1), 38-42.
- [Kiyota '93] Kiyota, Y., T. Nakamura, T. Inada, A. Kuranouchi und Y. Hirano (1993). *Characteristics of Shallow Boron-Doped Layers in Si by Rapid Vapor-Phase Direct Doping*. J. Electrochem. Soc. **140** (4), 1117-1121.
- [Kiyota '94] Kiyota, Y., T. Nakamura, K. Muraki und T. Inada (1994). *Phosphorus Direct Doping from Vapor Phase into Silicon for Shallow Junctions*. J. Electrochem. Soc. **141** (8), 2241-2244.
- [Kiyota '96] Kiyota, Y., T. Nakamura, S. Suzuki und T. Inada (1996). *Shallow p-Type Layers in Si by Rapid Vapor-Phase Doping for High-Speed Bipolar and MOS Applications*. IEICE Trans. Electron. **E79-C** (4), 554-559.
- [Koh '02] Koh, K. W., H. J. Oh, H. Choi, H. Kurino und M. Koyanagi (2002). *A novel atomic doping technology for ultra-shallow junction of SOI-MOSFETs*. Mater. Sci. Eng., B **89** (1-3), 435-438.
- [Lerch '98] Lerch, W., W. Blerch und S. Yanagawa (1998). *Temperature measurement of wafers with varying multilayer structures during rapid thermal annealing*. IEEE Transactions on Semiconductor Manufacturing **11** (4), 598-606.
- [Lerch '05] Lerch, W., S. Paul, et al. (2005). *Advanced activation of ultra-shallow junctions using flash-assisted RTP*. Mater. Sci. Eng., B **124-125** (SUPPL.), 24-31.
- [Li '03] Li, H.-J., T. Rhoad, P. Zeitzoff, R. Tichy, L. Larson und S. Banerjee (2003). *B diffusion in low energy B/BF₂ implants with pre-amorphization of different species*. Mater. Res. Soc. Symp. Proc. **765**, 161-166.
- [Lodders '99] Lodders, K. (1999). *Revised Thermochemical Properties of Phosphinidene (PH), Phosphine (PH₃), Phosphorus Nitride (PN), and Magnesium Phosphate (Mg₃P₂O₈)*. J. Phys. Chem. Ref. Data **28** (6), 1705-1712.
- [Lorenz '99] Lorenz, L., G. Deboy, A. Knapp und M. Maerz (1999). *CoolMOS - a new milestone in high voltage power MOS*. IEEE International Symposium on Power Semiconductor Devices and ICs (ISPSD), 3-10.
- [Madelung '91] Madelung, O. (1991). *Semiconductors : group IV elements and III-V compounds*. Berlin [u.a.], Springer.
- [Makris '73] Makris, J. S. und B. J. Masters (1973). *Phosphorus Isoconcentration Diffusion Studies in Silicon*. J. Electrochem. Soc. **120** (9), 1252-1255.
- [Masetti '83] Masetti, G., M. Severi und S. Solmi (1983). *Modeling of Carrier Mobility Against Carrier Concentration in Arsenic-, Phosphorus-, and Boron-Doped Silicon*. IEEE Trans. Electron Devices **ED-30** (7), 764-769.
- [Miotto '01a] Miotto, R., G. P. Srivastava und A. C. Ferraz (2001a). *Dissociative adsorption of PH₃ on the Si(001) surface*. Phys. Rev. B: Condens. Matter **63**, 125321 (9 pages).
- [Miotto '01b] Miotto, R., G. P. Srivastava, R. H. Miwa und A. C. Ferraz (2001b). *A comparative study of dissociative adsorption of NH₃, PH₃, and AsH₃ on Si(001)-(2×1)*. J. Chem. Phys. **114** (21), 9549-9556.
- [Moffatt '00] Moffatt, S., P. L. F. Hemment, S. Whelan und D. G. Armour (2000). *Silicon damage studies due to ultra-low-energy ion implantation with heavy species and rapid thermal annealing*. Mater. Sci. Semicond. Process. **3** (4), 291-296.
- [Moore '65] Moore, G. E. (1965). *Cramming more components onto integrated circuits*. Electronics **38** (8), 114-117.

- [Moore '75] Moore, G. E. (1975). *Progress in digital integrated circuit*. IEDM Techn. Dig., 11.
- [Morin '54] Morin, F. J. und J. P. Maita (1954). *Electrical Properties of Silicon Containing Arsenic and Boron*. Phys. Rev. **96**, 28.
- [Nakahara '96] Nakahara, Y., K. Takeuchi, T. Tatsumi, Y. Ochiai, S. Manako, S. Samukawa und A. Furukawa (1996). *Ultra-shallow in-situ-doped raised source/drain structure for sub-tenth micron CMOS*. Digest of Technical Papers - Symposium on VLSI Technology, 174-175.
- [Nenyei '93] Nenyei, Z., H. Walk und T. Knarr (1993). *Defect-Guarded Rapid Thermal Processing*. J. Electrochem. Soc. **140** (6), 1782-1733.
- [Oberbeck '03] Oberbeck, L., N. J. Curson, T. Hallam, M. Y. Simmons, R. G. Clark und G. Bilger (2003). *Measurement of phosphorus segregation in silicon at the atomic-scale using STM*. arXiv.org: cond-mat/0307495.
- [Oh '04] Oh, J., K. Im, C.-G. Ahn, J.-H. Yang, W.-j. Cho, S. Lee und K. Park (2004). *Ultra shallow and abrupt n^+ - p junction formations on silicon-on-insulator by solid phase diffusion of arsenic from spin-on-dopant for sub 50 nm Si metal-oxide-semiconductor devices*. Mater. Sci. Eng., B **110** (2), 185-189.
- [Olesinsky '86] Olesinsky, R. W., N. Kanani und G. J. Abbaschian (1986). Bull. Alloy Phase Diagrams **6**, 130.
- [Olsen '89] Olsen, J. E. und F. Shimura (1989). *Infrared analysis of film growth on the silicon surface in room air*. J. Vac. Sci. Technol., A **7** (6), 3275-3278.
- [Ono '95] Ono, M., M. Saito, T. Yoshitomi, C. Fiegna, T. Ohguro, H. S. Momose und H. Iwai (1995). *Fabrication of sub-50-nm gate length n-metal-oxide-semiconductor field effect transistors and their electrical characteristics*. J. Vac. Sci. Technol., B **13** (4), 1740-1743.
- [Plummer '00] Plummer, J. D., M. Deal und P. B. Griffin (2000). *Silicon VLSI technology: fundamentals, practice and modeling*. Upper Saddle River, NJ, Prentice Hall.
- [Privitera '02] Privitera, V. (2002). *Ultra-low energy ion implantation of boron for future silicon devices*. Current Opinion in Solid State and Materials Science **6** (1), 55-65.
- [Ransom '94] Ransom, C., T. N. Jackson, J. F. DeGelormo, C. Zeller, D. E. Kotecki, C. Graitmann, D. K. Sadana und J. Benedict (1994). *Shallow n^+ Junctions in Silicon by Arsenic Gas-Phase Doping*. J. Electrochem. Soc. **141** (5), 1378-1381.
- [Rubin '03] Rubin, L. und J. Poate (2003). *Ion implantation in silicon technology*. Industrial Physicist **9** (3), 12-15.
- [Ruffell '05] Ruffell, S., I. V. Mitchell und P. J. Simpson (2005). *Annealing behavior of low-energy ion-implanted phosphorus in silicon*. J. Appl. Phys. **97** (12), 123518.
- [Saga '97] Saga, K. und T. Hattori (1997). *Acceleration of organic contaminant adsorption onto silicon surfaces in the presence of residual fluorine*. J. Electrochem. Soc. **144** (9), L250-L-252.
- [Saito '93] Saito, M., T. Yoshitomi, et al. (1993). *P-MOSFET's with ultra-shallow solid-phase-diffused drain structure produced by diffusion from BSG gate-sidewall*. IEEE Trans. Electron Devices **40** (12), 2264-2272.
- [Sato '98] Sato, T., I. Mizushima, N. Aoki und Y. Tsunashima (1998). *Dominant factor for the concentration of phosphorus introduced by vapor phase doping (VPD)*. Jpn. J. Appl. Phys., Part 1 **37** (3B), 1162-1165.

- [Sato '97] Sato, Y., K. Imai und N. Yabumoto (1997). *Phosphorus Pileup and Sublimation at the Silicon Surface*. J. Electrochem. Soc. **144** (7), 2548-2551.
- [Sato '93] Sato, Y., M. Watanabe und K. Imai (1993). *Characterization of Phosphorus Pileup at the SiO₂/Si Interface*. J. Electrochem. Soc. **140** (9), 2679-2682.
- [Schroder '98] Schroder, D. K. (1998). *Semiconductor material and device characterization*. New York, Wiley.
- [Shao '03] Shao, L., J. Liu, Q. Y. Chen und W.-K. Chu (2003). *Boron diffusion in silicon: the anomalies and control by point defect engineering*. Mater. Sci. Eng., R **42** (3-4), 65-114.
- [Shauly '04] Shauly, E. N. und S. Lachman-Shalem (2004). *Activation improvement of ion implanted boron in silicon through fluorine co-implantation*. J. Vac. Sci. Technol., B **22** (2), 592-596.
- [Shimamune '04] Shimamune, Y., M. Sakuraba, J. Murota und B. Tillack (2004). *Formation of heavily P-doped Si epitaxial film on Si(100) by multiple atomic-layer doping technique*. Appl. Surf. Sci. **224** (1-4), 202-205.
- [SILVACO '05] SILVACO (2005). ATHENA User's Manual, 2d process simulation software, Version 5.10.0.R.
- [Silvestri '04] Silvestri, H. H., I. D. Sharp, E. E. Haller, H. A. Bracht, J. Hansen und A. Nylandsted-Larsen (2004). *Simultaneous phosphorus and Si self-diffusion in extrinsic, isotopically controlled silicon heterostructures*. Mater. Res. Soc. Symp. Proc., **810**, 77-83.
- [Skinner '06] Skinner, W., J. Hautala, M. Gwinn, Y. Shao und T. Tetreault (2006). *Infusion doping advances manufacturing*. European Semiconductor (Feb), 19-20.
- [Solmi '96] Solmi, S., A. Parisini, R. Angelucci, A. Armigliato, D. Nobili und L. Moro (1996). *Dopant and carrier concentration in Si in equilibrium with monoclinic SiP precipitates*. Phys. Rev. B: Condens. Matter **53** (12), 7836-7841.
- [Song '00] Song, Y.-H., K.-T. Park, H. Kurino und M. Koyanagi (2000). *Ultrashallow junction formation by rapid thermal annealing of arsenic-adsorbed layer*. Jpn. J. Appl. Phys., Part 1 **39** (1), 26-30.
- [Souza '88] Souza, J. P. d., C. M. Hasenack und J. E. Swart (1988). *The doping of silicon with boron by rapid thermal processing*. Semicond. Sci. Technol. **3**, 277-280.
- [Srivastava '02] Srivastava, G. P. (2002). *Surface passivation by dissociative molecular adsorption*. Vacuum **67** (1), 11-20.
- [Suni '98] Suni, I. I. (1998). *Ellipsometric Thickness Measurements of Ultrathin Silicon Oxides Formed in Aqueous Solution*. Jpn. J. Appl. Phys., Part 2 **37** (6B), L712-L714.
- [Taylor '04] Taylor, W. J., J. Jiang, et al. (2004). *Materials challenges for CMOS junctions*. Mater. Res. Soc. Symp. Proc., **810**, 3-14.
- [Thakur '91] Thakur, R. P. S., R. Singh, A. J. Nelson, H. S. Ullal, J. Chaudhuri und V. Gondhalekar (1991). *Comparative study of phosphosilicate glass on (100) silicon by furnace and rapid isothermal annealing*. J. Appl. Phys. **69** (1), 367-371.
- [Theunissen '99] Theunissen, M. J. J., C. E. Timmering, J. G. v. Berkum und Y. J. Mergler (1999). *Boron vapor phase doping of silicon for bipolar device applications*. Jpn. J. Appl. Phys., Part 1 **38** (10), 5805-5814.
- [Tillack '98] Tillack, B. (1998). *Atomic control of doping during SiGe epitaxy*. Thin Solid Films **318**, 1-5.

- [Tillack '00] Tillack, B., B. Heinemann und D. Knoll (2000). *Atomic layer doping of SiGe – fundamentals and device applications*. Thin Solid Films **369** (1-2), 189-194.
- [Tillmann '95] Tillmann, A. (1995). *Optimization of Process Parameters and Temperature Uniformity on Wafers for Rapid Thermal Processing*. Mater. Res. Soc. Symp. Proc., **387**, San Francisco, 3-14.
- [Tompkins '99] Tompkins, H. G. und W. A. McGahan (1999). *Spectroscopic ellipsometry and reflectometry : a user's guide*. New York, NY [u.a.], Wiley.
- [Tsukidate '99] Tsukidate, Y. und M. Suemitsu (1999). *Saturated adsorption of PH₃ on Si(100):P and its application to digital control of phosphorus coverage on Si(100) surface*. Appl. Surf. Sci. **151** (1-2), 148-152.
- [Uchino '04] Uchino, T., P. Ashburn, Y. Kiyota und T. Shiba (2004). *A CMOS-compatible rapid vapor-phase doping process for CMOS scaling*. IEEE Trans. Electron Devices **51** (1), 14-19.
- [Uchino '01] Uchino, T., A. Miyauchi und T. Shiba (2001). *MOSFETs with ultrashallow junction and minimum drain area formed by using solid-phase diffusion from SiGe*. IEEE Trans. Electron Devices **48** (7), 1406-1411.
- [Uematsu '97] Uematsu, M. (1997). *Simulation of boron, phosphorus, and arsenic diffusion in silicon based on an integrated diffusion model, and the anomalous phosphorus diffusion mechanism*. J. Appl. Phys. **82** (5), 2228-2246.
- [Usami '92] Usami, A., M. Ando, M. Tsunekane und T. Wada (1992). *Shallow-junction formation on silicon by rapid thermal diffusion of impurities from a spin-on source*. IEEE Trans. Electron Devices **39** (1), 105-110.
- [Vandervorst '84] Vandervorst, W., H. E. Maes und R. F. D. Keersmaecker (1984). *Secondary ion mass spectrometry: Depth profiling of shallow As implants in silicon and silicon dioxide*. J. Appl. Phys. **56** (5), 1425-1433.
- [Vandervorst '87] Vandervorst, W. und F. R. Shepherd (1987). *Secondary ion mass spectrometry profiling of shallow, implanted layers using quadrupole and magnetic sector instruments*. J. Vac. Sci. Technol., A **5** (3), 313-320.
- [Vlad '99] Vlad, M. O., G. Cerofolini und J. Ross (1999). *Temporal Logarithmic Oscillations in Self-Similar Multilayer Aggregation: Shlesinger-Hughes Renormalization with Application to the Tunnel-Assisted Wet Oxidation of Silicon*. J. Phys. Chem. A **103** (25), 4798-4807.
- [Vuong '00] Vuong, H.-H., C. S. Rafferty, S. A. Eshraghi, J. Ning, J. R. McMacken, S. Chaudhry, J. McKinley und F. A. Stevie (2000). *Dopant dose loss at the Si-SiO₂ interface*. J. Vac. Sci. Technol., B **18** (1), 428-434.
- [Watanabe '98] Watanabe, H., K. Kato, T. Uda, K. Fujita, M. Ichikawa, T. Kawamura und K. Terakura (1998). *Kinetics of Initial Layer-by-Layer Oxidation of Si(001) Surfaces*. Phys. Rev. Lett. **80** (2), 345-348.
- [Watanabe '95] Watanabe, S. und Y. Sugita (1995). *The role of dissolved oxygen in hot water during dissolving oxides and terminating silicon surfaces with hydrogen*. Surf. Sci. **327** (1-2), 1-8.
- [Whelan '02] Whelan, S., V. Privitera, et al. (2002). *Redistribution and electrical activation of ultra-low energy implanted boron in silicon following laser annealing*. J. Vac. Sci. Technol., B **20** (2), 644-649.

- [Wilson '89] Wilson, R. G., F. A. Stevie und C. W. Magee (1989). *Secondary ion mass spectrometry: a practical handbook for depth profiling and bulk impurity analysis*. New York u.a., John Wiley & Sons.
- [Wittel '95] Wittel, F. und S. Dunham (1995). *Diffusion of phosphorus in arsenic and boron doped silicon*. Appl. Phys. Lett. **66** (11), 1415-1417.
- [Wittmaack '84] Wittmaack, K. (1984). *Beam-induced broadening effects in sputter depth profiling*. Vacuum **34** (1-2), 119-137.
- [Wittmaack '85] Wittmaack, K. (1985). *Assessment of the Relative Contribution of Atomic Mixing and Selective Sputtering to Beam Induced Broadening in SIMS Depth Profiling*. Nucl. Instrum. Methods Phys. Res., Sect. B **7-8** (2), 750-754.
- [Wong '02] Wong, H.-S. P. (2002). *Beyond the conventional transistor*. IBM J. Res. & Dev. **46** (3-3), 133-168.
- [Wright '82] Wright, S. L. und H. Kroemer (1982). *Operational aspects of a gallium phosphide source of P₂ vapor in molecular beam epitaxy*. J. Vac. Sci. Technol. **20** (2), 143-148.
- [Yoshida '02] Yoshida, M. und S. Tanaka (2002). *Simulation of Phosphorus Diffusion Profiles with Different Phosphorus Surface Concentration at the Same Diffusion Temperature in Silicon*. Jpn. J. Appl. Phys., Part 1 **41** (9), 5493-5502.
- [Yu '84] Yu, M. L. und B. S. Meyerson (1984). *The adsorption of PH₃ on Si(100) and its effect on the coadsorption of SiH₄*. J. Vac. Sci. Technol., A **2** (2), 446-449.