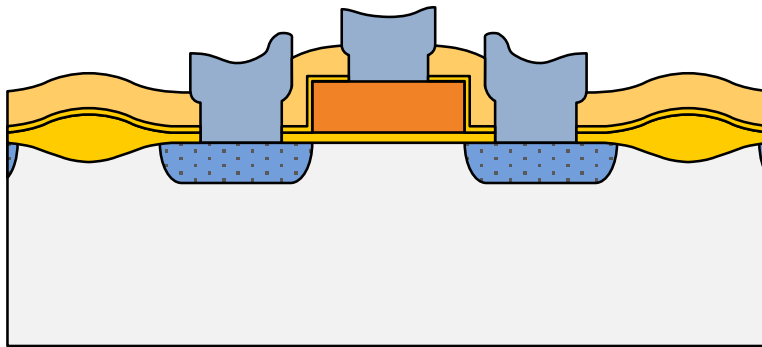


Halbleitertechnologie

von A bis Z



www.halbleiter.org

Inhaltsverzeichnis

Abbildungsverzeichnis	VI
Tabellenverzeichnis	X
1 Grundlagen	1
1.1 Der Atombau	1
1.1.1 Das Atommodell	1
1.1.2 Größenordnungen	2
1.2 Das Periodensystem der Elemente	2
1.2.1 Die Elemente	2
1.2.2 Das Periodensystem	3
1.3 Chemische Bindungen	5
1.3.1 Chemische Bindungen	5
1.3.2 Die Atombindung	5
1.3.3 Die Ionenbindung	6
1.3.4 Die Metallbindung	7
1.4 Edelgase	8
1.5 Leiter - Nichtleiter - Halbleiter	9
1.5.1 Leiter	9
1.5.2 Nichtleiter	10
1.5.3 Halbleiter	10
1.5.4 Das Bändermodell	11
1.6 Dotieren: n- und p-Halbleiter	14
1.6.1 Dotieren	14
1.6.2 n-Dotierung	15
1.6.3 p-Dotierung	16
1.6.4 Bänderschema bei dotierten Halbleitern	17

1.7	Der p-n-Übergang	17
1.7.1	p-n-Übergang ohne angelegte Spannung	17
1.7.2	p-n-Übergang mit angelegter Spannung	18
1.8	Aufbau eines n-Kanal-FET	19
1.8.1	Allgemeiner Aufbau	19
1.8.2	Aufbau eines n-Kanal-FET	20
1.8.3	Funktionsweise	25
1.9	Aufbau eines Bipolartransistors	26
1.9.1	Allgemeiner Aufbau	26
1.9.2	Aufbau eines npn-Bipolartransistors	27
1.9.3	Funktionsweise	29
1.10	Aufbau eines FinFET	30
1.10.1	Allgemeiner Aufbau und Funktionsweise	30
1.10.2	Aufbau eines FinFET im Bulkprozess	31
2	Waferherstellung	34
2.1	Silicium	34
2.2	Herstellung des Rohsiliciums	35
2.2.1	Herstellung von Rohsilicium	35
2.2.2	Reinigung von Rohsilicium	35
2.2.3	Zonenreinigung	36
2.3	Herstellung des Einkristalls	37
2.3.1	Der Einkristall	37
2.3.2	Kristallziehverfahren nach Czochralski	38
2.3.3	Tiegelfreies Zonenziehen	39
2.4	Entstehung der Wafer	40
2.4.1	Wafervereinzelung und Oberflächenveredelung	40
2.4.2	Geschichtliche Entwicklung der Wafergröße	42
2.5	Dotiertechniken	44
2.5.1	Begriff	44
2.5.2	Diffusion	44
2.5.3	Diffusionsverfahren	45
2.5.4	Ionenimplantation	47

3	Oxidation	51
3.1	Übersicht	51
3.1.1	Anwendungen	51
3.1.2	Eigenschaften von Oxid	51
3.2	Erzeugung von Oxidschichten	52
3.2.1	Thermische Oxidation	52
3.2.2	Oxidation durch Abscheidung	55
3.3	Die LOCOS-Technik	56
3.3.1	Höchstintegration auf Chips	56
3.3.2	Der Vogelschnabel	57
3.4	Schichtdickenmessung	59
3.4.1	Messtechnik	59
3.4.2	Interferometrie	59
3.4.3	Ellipsometrie	60
3.4.4	Beurteilung der Messung	61
4	Abscheidung	63
4.1	Plasma, der 4. Aggregatzustand	63
4.1.1	Plasmazustand	63
4.1.2	Plasmaerzeugung	64
4.2	CVD-Verfahren	65
4.2.1	Siliciumgasphasenepitaxie	65
4.2.2	CVD-Verfahren: Chemical Vapor Deposition	67
4.2.3	APCVD: Atmospheric Pressure CVD	68
4.2.4	LPCVD: Low Pressure CVD	69
4.2.5	PECVD: Plasma Enhanced CVD	70
4.2.6	ALD: Atomic Layer Deposition	71
4.3	PVD-Verfahren	73
4.3.1	MBE: Molekularstrahlepitaxie	73
4.3.2	Aufdampfen	74
4.3.3	Sputtern: Kathodenzerstäubung	75
5	Metallisierung	78
5.1	Anforderungen an die Metallisierung	78
5.2	Aluminiumtechnologie	79
5.2.1	Aluminium und Aluminiumlegierungen	79

5.2.2	Siliciumdiffusion	79
5.2.3	Elektromigration	81
5.2.4	Hillockwachstum	81
5.3	Kupfertechnologie	82
5.3.1	Kupfertechnologie	82
5.3.2	Damascene-Verfahren	83
5.3.3	Low-k-Technologie	87
5.4	Der Metall-Halbleiter-Kontakt	90
5.4.1	Der Metall-Halbleiter-Kontakt	90
5.4.2	Bändermodell eines p-n-Übergangs	94
5.5	Mehrlagenverdrahtung	95
5.5.1	Mehrlagenverdrahtung	95
5.5.2	BPSG-Reflow	95
5.5.3	Reflowrückätzen	95
5.5.4	Chemisch-Mechanisches Polieren	96
5.5.5	Kontaktierung der Metallisierungsebene	98
6	Lithografie	100
6.1	Belichten und Belacken	100
6.1.1	Übersicht	100
6.1.2	Aufbringen eines Haftvermittlers	100
6.1.3	Belacken	101
6.1.4	Belichtung	102
6.1.5	Eingesetzte Belichtungsverfahren	104
6.2	Belichtungsverfahren	104
6.2.1	Übersicht	104
6.2.2	Kontaktbelichtung	105
6.2.3	Abstandsbelichtung	106
6.2.4	Reduzierende Projektionsbelichtung	106
6.2.5	Elektronenstrahlolithografie	107
6.2.6	Röntgenstrahlolithografie	108
6.2.7	Weitere Verfahren	109
6.3	Der Fotolack	110
6.3.1	Lacktechnik	110
6.3.2	Chemische Zusammensetzung	111

6.4	Entwickeln und Kontrolle	111
6.4.1	Entwickeln	111
6.4.2	Lackkontrolle	112
6.4.3	Lackentfernen	113
6.5	Maskentechnik	114
6.5.1	Maskentechnik	114
6.5.2	Schritte bei der Maskenherstellung	115
6.5.3	Maskentypen	116
6.5.4	Belichtungsverfahren der nächsten Generation	118
7	Nasschemie	120
7.1	Ätztechnik allgemein	120
7.2	Nassätzen	121
7.2.1	Prinzip	121
7.2.2	Anforderung	121
7.2.3	Tauchätzung	122
7.2.4	Sprühätzung	123
7.2.5	Anisotrope Siliciumätzung	123
7.2.6	Ätzlösungen für isotrope Ätzungen	124
7.3	Scheibenreinigung	125
7.3.1	Der Reinraum	125
7.3.2	Arten der Verunreinigung	127
7.3.3	Mikroskopische Verunreinigungen	127
7.3.4	Molekulare Verunreinigungen	128
7.3.5	Alkalische und metallische Verunreinigungen	128
7.3.6	Reinigungstechniken	129
8	Trockenätzen	131
8.1	Übersicht	131
8.1.1	Allgemein	131
8.1.2	Wichtige Größen beim Trockenätzen	131
8.1.3	Trockenätzverfahren	132
8.2	Trockenätzverfahren	132
8.2.1	Ionenstrahlätzen	132
8.2.2	Plasmaätzen	134
8.2.3	Reaktives Ionenätzen	135

Abbildungsverzeichnis

1.1	Stark vereinfachte Darstellung eines Neonatoms	1
1.2	Atombausteine	2
1.3	Atome die eine große Bedeutung in der Halbleiterindustrie haben	3
1.4	Periodensystem der Elemente	4
1.5	Atombindung am Beispiel von Silan	6
1.6	Prinzip der Ionenbindung am Beispiel von Natriumchlorid	7
1.7	Metallbindung	8
1.8	Edelgase	9
1.9	Metallbindung: Feste Atomrümpfe und freie Valenzelektronen	9
1.10	Ausschnitt aus einem Siliciumkristall	11
1.11	Aufspaltung von Energieniveaus durch Wechselwirkung von Atomen .	12
1.12	Energiebänder durch in Wechselwirkung stehende Atome	12
1.13	Das Bändermodell	14
1.14	n-Dotierung mit Phosphor	15
1.15	p-Dotierung mit Bor	16
1.16	Bändermodell von n- und p-Dotierten Halbleitern	17
1.17	p-n-Übergang ohne angelegte Spannung	18
1.18	p-n-Übergang mit angeschlossener Spannungsquelle	19
1.19	Anreicherungs-Feldeffekttransistor	26
1.20	SBC-Bipolartransistor	30
1.21	FinFET im Bulk-Prozess	33
1.22	FinFET im SOI-Prozess	33
2.1	Darstellung der Zonenreinigung	37
2.2	Kristallorientierungen	37
2.3	Kristallziehen nach Czochralski	38
2.4	Tiegelfreies Zonenziehen	40
2.5	Verschiedene Wafertypen	41

2.6	Innenloch- und Drahtsäge	41
2.7	Kantenverrundung	42
2.8	Wafergrößen in der Übersicht	43
2.9	Diffusionsvorgänge im Kristallgitter	45
2.10	Diffusion mit Oxidmaske	46
2.11	Darstellung einer Implantationsanlage	48
2.12	Channelingeffekt im Einkristallgitter	49
3.1	Darstellung eines Oxidationsofens	52
3.2	Aufwuchsverhalten von Oxid auf Silicium	55
3.3	Schichtaufbau vor dem LOCOS-Prozess	57
3.4	LOCOS-Struktur nach der Oxidation	58
3.5	Anwendungsbeispiel der LOCOS-Technik zur seitlichen Isolation	59
3.6	Prinzip der Interferenz, Überlagerung von Lichtwellen	60
3.7	Ellipsometrie	61
3.8	Vergleich von Messung und Simulation	62
4.1	Prinzip der Plasmaerzeugung	65
4.2	Darstellung eines Barrelreaktors zur Epitaxieabscheidung	67
4.3	Darstellung eines Horizontalreaktors zur APCVD-Abscheidung	69
4.4	Darstellung einer LPCVD-Abscheidungsanlage für TEOS-Schichten	70
4.5	Darstellung einer PECVD-Anlage und einer Horde mit Wafern	71
4.6	Einlass von TMA in die Prozesskammer	72
4.7	Anlagerung von TMA an der Waferoberfläche	72
4.8	Einlass von Wasserdampf in die Prozesskammer	73
4.9	Abgesättigte Oberfläche vor dem nächsten Zyklus	73
4.10	Schematische Darstellung einer MBE-Anlage	74
4.11	Schematische Darstellung einer Aufdampfanlage	75
4.12	Schematische Darstellung einer Sputteranlage	77
5.1	Spikebildung	79
5.2	Kontakimplantation	80
5.3	Barriereschicht zwischen Aluminium und Substrat	80
5.4	Kantenverrundung zur konformen Aluminiumabscheidung	81
5.5	Prinzipieller Damascene-Schichtaufbau vor der Strukturierung	84
5.6	Grabenätzung und VIA-Lackmaske	84
5.7	Finale Struktur nach dem Ätzen und der Kupferabscheidung	85

5.8	Damascene-Struktur nach Metallisierung und Planarisierung	85
5.9	Damascene-Schichtaufbau nach dem Kontaktlochätzen bei VFTL	86
5.10	Fototechnik und Ätzung der Gräben	86
5.11	Damascene-Struktur nach Metallisierung und Planarisierung	87
5.12	Klassischer Plattenkondensator und parasitäre Kapazität	87
5.13	k-Wert in Abhängigkeit von der Porosität	89
5.14	Ferminiveau in Metallen	91
5.15	Ferminiveau in dotierten Halbleitern	91
5.16	Bändermodell vor dem Kontakt von Metall und n-Silicium	92
5.17	Bändermodell nach dem Kontakt von Metall und n-Silicium	92
5.18	Bändermodell nach n^+ -Dotierung	93
5.19	Bändermodell nach dem Kontakt von Metall und p-Silicium	94
5.20	Bändermodell am Übergang p- und n-dotierter Siliciumkristalle	94
5.21	Prozessschritte beim Reflowrückätzen	96
5.22	Schematische Darstellung einer CMP-Anlage	97
5.23	CMP-Prozess zur Grabenisolation	98
6.1	Oberflächenmodifikation mit HMDS	101
6.2	Schleuderbelackung	102
6.3	Prinzip der optischen Fotolithografie	102
6.4	Fehler bei der Belichtung	103
6.5	Kontaktbelichtung	105
6.6	Abstandsbelichtung	106
6.7	Step-and-Repeat-Belichtung	107
6.8	Elektronenstrahlolithografie	108
6.9	Röntgenstrahlolithografie	109
6.10	Darstellung der belichteten Lackarten nach dem Entwickeln	112
6.11	Darstellung von Justiermarken	113
6.12	Prinzipieller Schichtaufbau einer Chrom-on-Glass-Maske	115
6.13	Entwickeln der Lackschicht	115
6.14	Ätzen der Chromschicht	115
6.15	Lackentfernen	115
6.16	Anbringen des Pellicles	116
6.17	Intensitätsprofil einer Chrom On Glass-Mask	116
6.18	Intensitätsprofil einer Attenuated Phase Shift Mask, Prinzip der Phasenschiebung	117

6.19	Intensitätsprofil einer Alternating Phase Shift Mask	118
6.20	Zielstruktur auf dem Wafer und entsprechende Maskenstruktur	119
7.1	Isotroper und anisotroper Ätzvorgang	120
7.2	Tauchätzung	122
7.3	Sprühätzung	123
7.4	Reinraumklassen	126
7.5	Abschattung bei Ätz- und Implantationsprozessen durch Partikel	128
7.6	Partikeleinschluss	128
8.1	Darstellung eines Ionenstrahl-Ätzreaktors	133
8.2	Darstellung einer CDE-Ätzanlage	134
8.3	Bildung von Radikalen	135
8.4	Darstellung eines RIE-Reaktors in Hexodenbauform	136
8.5	Darstellung eines RIE-Reaktors in Parallelplattenbauweise	137
8.6	Seitenwandpassivierung beim RIE-Ätzen	138

Tabellenverzeichnis

1.1	Bedeutende Elemente in der Halbleitertechnologie	5
2.1	Typische Daten von Wafern	43
3.1	Vergleich der Aufwachsraten bei trockener und nasser Oxidation	54
5.1	Übersicht verschiedener organischer Siliciumoxide	90
8.1	Übersicht über Prozessgase zum Trockenätzen verschiedener Schichten	139

1 Grundlagen

1.1 Der Atombau

1.1.1 Das Atommodell

Ein Atom ist der kleinste chemisch nicht weiter teilbare Baustein der Materie. Je nach Atomart setzt es sich aus einer bestimmten Anzahl an Elektronen, Protonen und Neutronen zusammen. Die positiv geladenen Protonen bilden zusammen mit den Neutronen den Atomkern, den die Elektronen in bestimmten Abständen umkreisen. Ein in der Natur vorkommendes Atom ist elektrisch neutral, es befinden sich also genauso viele positive Protonen wie negativ geladene Elektronen im Atom. Die Anzahl der Neutronen kann dabei variieren. Das einfachste Atom ist das Wasserstoffatom, es besitzt ein Elektron, ein Proton und kein Neutron. Im nächsten Atom, dem Edelgas Helium, befinden sich zwei Elektronen, zwei Protonen und zwei Neutronen.

Nach dem Bohrschen Atommodell befinden sich die Elektronen auf sogenannten Schalen, welche unterschiedliche Energieniveaus repräsentieren und konzentrisch um den Atomkern angeordnet sind. Es gibt maximal sieben Schalen, die unterschiedlich viele Elektronen aufnehmen können, die Elektronen auf der äußersten Schale werden auch als Valenzelektronen bezeichnet.

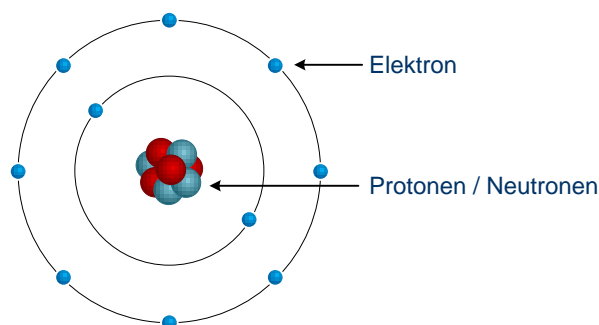


Abb. 1.1: Stark vereinfachte Darstellung eines Neonatoms

Das Bestreben aller Atome ist es, die äußerste Schale voll zu besetzen, mit acht Elektronen erreichen sie die so genannte Edelgaskonfiguration (auch Elektronenoktett). Elemente mit wenigen Außenelektronen geben dazu Elektronen ab, Elemente mit vielen Valenzelektronen nehmen zusätzliche Elektronen auf (Details im Kapitel chemische Bindungen).

1.1.2 Größenordnungen

Masse: Die Masse eines Atoms wird hauptsächlich vom Atomkern bestimmt, da die Massen von Protonen und Neutronen mit $1,67 \cdot 10^{-27}$ kg rund 1800 Mal größer sind, als die der Elektronen in der Atomhülle ($9,11 \cdot 10^{-31}$ kg).

Abmessungen: Der Durchmesser der Atomhülle beträgt 0,1-0,5 nm, der Durchmesser des Atomkerns ist noch einmal um den Faktor 100.000 geringer. Zur Veranschaulichung: wenn ein Stecknadelkopf in der Mitte eines Fußballfeldes den Atomkern darstellt, so entspricht die Entfernung zu den Eckfahnen dem Abstand, mit dem die Elektronen den Kern umkreisen.

Dichte: Im Kern eines Atoms sind Protonen und Neutronen extrem dicht gepackt. Würde man die Erde auf die gleiche Dichte komprimieren, so würde ihr Radius von ursprünglich 6.700.000 m auf nur noch 100 m sinken.




			
Teilchen	Proton	Neutron	Elektron
Ladung	+1	0	-1
Masse	$1,672 \cdot 10^{-27}$ kg	$1,675 \cdot 10^{-27}$ kg	$0,0009 \cdot 10^{-27}$ kg

Abb. 1.2: Atombausteine

1.2 Das Periodensystem der Elemente

1.2.1 Die Elemente

Ein Element besteht aus mehreren gleichen Atomen (das bedeutet, mit derselben Anzahl an Protonen = Kernladungszahl) und ist ein Stoff, der mit chemischen Mitteln

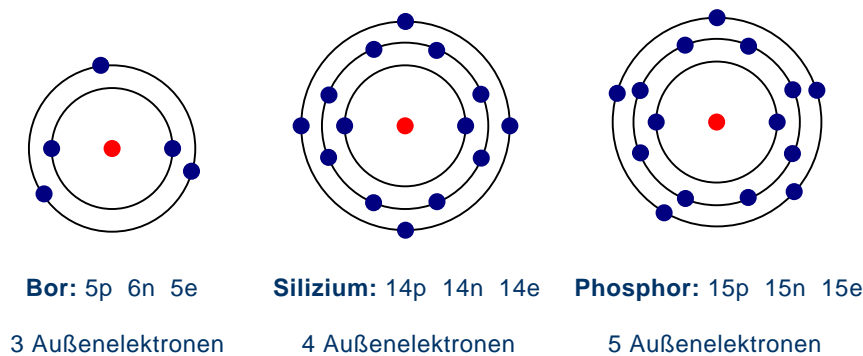


Abb. 1.3: Atome die eine große Bedeutung in der Halbleiterindustrie haben

nicht weiter zerlegt werden kann. Die Masse von Elementen wird nur durch die Anzahl von Protonen und Neutronen bestimmt, da die Elektronenmasse vernachlässigbar gering ist. Wasserstoff mit einem Proton und keinem Neutron hat die Massenzahl 1, das nächst schwerere Element, Helium, besitzt die Massenzahl 4 (2 Protonen + 2 Neutronen). Die Massenzahl gibt die Anzahl der Teilchen im Atomkern an. Multipliziert mit der atomaren Masseneinheit $u = 1,660 \cdot 10^{-27} \text{ kg}$ erhält man in etwa die Masse eines Atoms, für Helium $6,64 \cdot 10^{-27} \text{ kg}$.

Die Elemente, bis heute sind ca. 117 bekannt, werden meist mit den Anfangsbuchstaben ihrer lateinischen oder griechischen Namen benannt (Wasserstoff H von lat. hydrogenium, Lithium Li von gr. lithos).

1.2.2 Das Periodensystem

Das Periodensystem der Elemente (kurz PSE) stellt alle chemischen Elemente mit steigender Protonenanzahl (Kernladung, Ordnungszahl) und entsprechend ihrer chemischen Eigenschaften eingeteilt in Perioden sowie Haupt- und Nebengruppen dar.

Die Periode gibt dabei die Anzahl der Elektronenschalen an, die Hauptgruppe die Anzahl der Elektronen auf der äußersten Schale (1 bis 8 Elektronen). Gruppe 1 und 2 sowie 13-18 bilden die Hauptgruppen, die Gruppen 3 – 12 die Nebengruppen.

Das erste Element mit einer Schale (Periode 1) und einem Außenelektron (Gruppe 1) ist Wasserstoff H. Das nächste Element, Helium He, besitzt wie Wasserstoff nur eine Elektronenschale und befindet sich somit ebenfalls in Periode 1. Da die erste Schale mit zwei Elektronen bereits vollständig gefüllt ist, steht Helium nicht in Gruppe 2, sondern in Gruppe 18 (Gruppe der Edelgase).

Um weitere Elektronen aufnehmen zu können, muss eine neue Schale begonnen werden. Somit findet man Lithium in Gruppe 1, Periode 2 (zwei Elektronen auf der ersten Schale, ein Valenzelektron auf der zweiten Schale). Eine Schale kann maximal $2n^2$ Elektronen aufnehmen, wobei n für die Periode steht.

Nachdem in Gruppe 1 und 2 die ersten beiden Valenzelektronen auf der äußersten Schale besetzt wurden, werden ab der vierten Periode zunächst weiter innenliegende Schalen mit Elektronen vervollständigt, bevor die jeweils äußerste Schale der Gruppen 13 bis 18 vollständig mit Elektronen aufgefüllt wird.

		Hauptgr.		Nebengruppen									Hauptgruppen						
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Periode	1	H																	He
	2	Li	Be											B	C	N	O	F	Ne
	3	Na	Mg											Al	Si	P	S	Cl	Ar
	4	K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
	5	Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	Te	I	Xe
	6	Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
	7	Fr	Ra	Ac	Rf	Db	Sg	Bh	Hs	Mt	Ds	Rg	Uub	Uut	Uuq	Uup	Uuh	Uus	Uuo
				Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu		
				Th	Ph	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr		
<hr/>																			
Alkalimetalle				Metalle				Halogene				Lanthanoide							
Erdalkalimetalle				Halbmetalle				Edelgase				Actinoide							
Übergangsmetalle				Nichtmetalle															

Abb. 1.4: Periodensystem der Elemente

Elemente, die sich links im Periodensystem befinden, sind Metalle. Diese haben das Bestreben, Valenzelektronen abzugeben um so eine stabile Elektronenkonfiguration (die Edelgaskonfiguration) zu erreichen. Rechts im Periodensystem stehen die Nichtmetalle, die zum Erreichen der Edelgaskonfiguration zusätzliche Elektronen aufnehmen. Dazwischen befinden sich die Halbmetalle wie Silicium und Germanium.

Element	Teilchen	Eigenschaften, Anwendungen
B Bor	5p, 6n, 5e	3 Außenelektronen: Zur p-Dotierung von Silicium
N Stickstoff	7p, 7n, 7e	Stabiles N ₂ -Molekül: Schutz und Spülgas, Schutzschichten auf dem Wafer
O Sauerstoff	8p, 8n, 8e	Sehr reaktionsfreudig: Oxidation von Silicium, Isolationschichten (SiO ₂) u.a.
F Fluor	9p, 10n, 10e	Reaktionsfreudigstes Element: wird in Verbindung mit anderen Stoffen zum Ätzen verwendet (z.B. HF, CF ₄)
Si Silicium	14p, 14n, 14e	Grundmaterial in der Halbleitertechnik
P Phosphor	15p, 16n, 15e	5 Außenelektronen: Zur n-Dotierung von Silicium

Tab. 1.1: Bedeutende Elemente in der Halbleitertechnologie

1.3 Chemische Bindungen

1.3.1 Chemische Bindungen

Elektronen, die sich auf der äußersten Schale befinden, können sich vom Atom lösen (z.B. durch Zuführen von Energie in Form von Wärme) und mit anderen Atomen ausgetauscht werden. Verbindungen von mehreren Elementen nennt man Moleküle. Der Grund für das Bindungsbestreben ist das Erreichen der mit acht Elektronen voll besetzten äußersten Schale: die so genannte Edelgaskonfiguration. Stoffe, die die volle Außenschale erreicht haben, gehen keine Verbindungen ein (einige wenige Ausnahmen wie z.B. Xenon-Fluor-Verbindungen sind möglich). Dabei unterscheidet man hauptsächlich zwischen drei verschiedenen Arten von Bindungen, die im Folgenden näher erläutert werden.

1.3.2 Die Atombindung

Nichtmetalle gehen diese Verbindung ein um das Elektronenoktett zu vervollständigen. Zwei Fluoratom (je sieben Außenelektronen) können durch gegenseitigen Austausch eines Elektrons ihr Elektronenoktett auffüllen. Der Abstand zwischen den beiden Atomkernen repräsentiert einen Kompromiss zwischen der Anziehung von Atomkern und Bindungselektronen und der Abstoßung der beiden Atomkerne. Der Grund für Atombindungen ist, dass die Natur immer bestrebt ist, den energetisch niedrigsten Zustand herzustellen. Da den Elektronen durch den Zusammenschluss mehrerer Atome zu einem Molekül mehr Raum zur Verfügung steht, was einer geringeren Energie

entspricht, kommt es überhaupt erst zur Atombindung.

Aus dem Bestreben zum Erreichen der voll besetzten Außenschale ergibt sich, dass Fluoratome elementar niemals als einzelnes Atom, sondern immer als Fluormolekül F_2 auftreten. Ebenso Stickstoff N_2 , Sauerstoff O_2 , Chlor Cl_2 , Brom Br_2 und Jod I_2 . Auf Grund der Elektronenpaare nennt man diese Bindung auch Elektronenpaarbindung oder kovalente Bindung.

Darstellung der Atombindung am Beispiel von Silan (SiH_4 ; beim Silicium ist nur die äußerste Schale abgebildet). Das Siliciumatom hat die volle Achterschale, die Wasserstoffatome die, bereits mit zwei Elektronen, voll besetzte erste Schale.

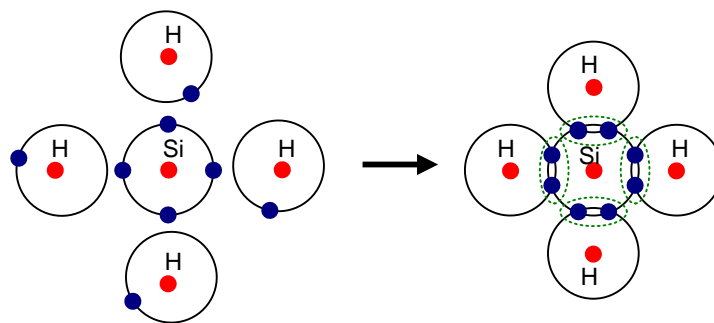


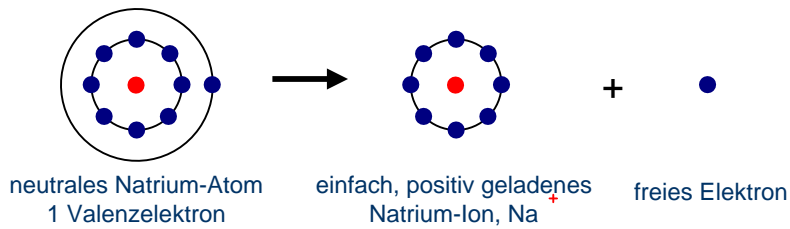
Abb. 1.5: Atombindung am Beispiel von Silan

1.3.3 Die Ionenbindung

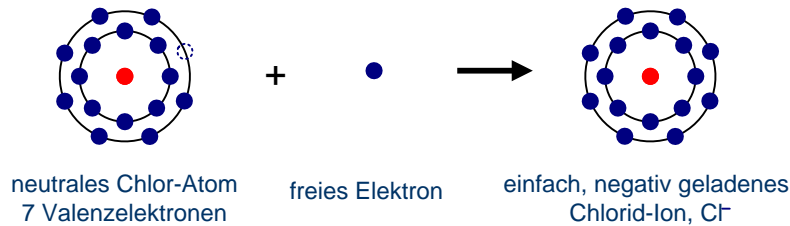
Ionenbindungen entstehen durch den Zusammenschluss von Metallen und Nichtmetallen. Während Metalle Elektronen abgeben wollen um eine vollständig gefüllte Außenschale zu erreichen, möchten Nichtmetalle zusätzliche Elektronen aufnehmen. Ein Beispiel für eine Ionenbindung ist Natriumchlorid $NaCl$ (Kochsalz).

Das Natriumatom gibt sein Valenzelektron ab (damit besitzt es mehr Protonen als Elektronen und ist positiv geladen), Chlor nimmt ein Elektron auf und ist damit einfach negativ geladen. Durch die unterschiedlichen Ladungen ziehen sich die zwei Atome an. Ein geladenes Atom bezeichnet man als Ion, dabei unterscheidet man zwischen Kation (positive Ladung) und Anion (negative Ladung).

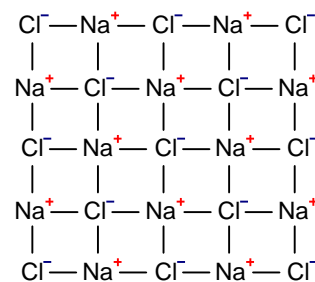
Da die Atome immer in einer sehr hohen Anzahl auftreten, richten sie sich dabei Dank der Anziehungs- und Abstoßungskräfte zu einem gleichmäßigen Ionengitter aus. Stoffe, die im festen Zustand ein solches Gitter bilden, bezeichnet man als Salze.



(a) Natrium als positives Kation



(b) Chlor als negatives Anion



(c) NaCl-Ionengitter

Abb. 1.6: Prinzip der Ionenbindung am Beispiel von Natriumchlorid

1.3.4 Die Metallbindung

Metalle gehen diese Bindung ein, um die stabile Edelgaskonfiguration zu erreichen. Dazu gibt jedes Metallatom seine Außenelektronen ab: es entstehen positiv geladene Metallionen (Atomrümpfe) und freie Elektronen zwischen denen starke Anziehungskräfte herrschen. Die Metallionen stoßen sich untereinander ebenso ab wie die Elektronen.

Da die Anziehungs- und Abstoßungskräfte in alle Richtungen des Raumes wirken ordnen sich die Atomrümpfe zu einem regelmäßigen Gitter an. In den Zwischenräumen befinden sich die frei beweglichen Elektronen als so genanntes Elektronengas, dieses

hält die positiven Metallionen zusammen. Auf Grund der frei beweglichen Elektronen leiten Metalle den elektrischen Strom sehr gut.

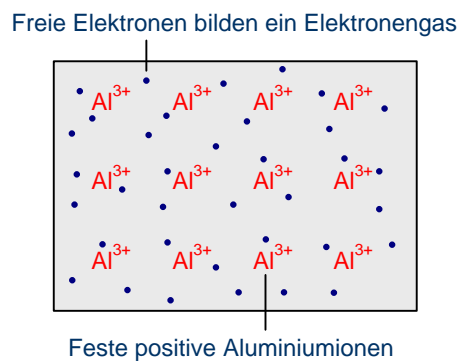


Abb. 1.7: Metallbindung

Die physikalischen und chemischen Eigenschaften der Verbindungen sind von der Bindungsart abhängig. So bedeuten stärkere Anziehungskräfte höhere Schmelz- und Siedepunkte, die Anzahl der freien Elektronen beeinflusst die Leitfähigkeit und auch der Ablauf von Reaktionen verschiedener Stoffe miteinander wird durch die Bindungsart bestimmt.

1.4 Edelgase

Edelgase sind die Elemente in der achten Hauptgruppe im Periodensystem: Helium, Neon, Argon, Krypton, Xenon, Radon (v.o.n.u.).

Die Besonderheit dieser Elemente ist, dass sie auf der äußersten Schale acht Elektronen besitzen. Dieses Elektronenoktett repräsentiert einen energetisch sehr stabilen Zustand, den alle Elemente erreichen wollen. Im Kapitel Bindungen wird erläutert, wie die Elemente die volle Achterschale erreichen. Auf Grund des stabilen Zustands der Edelgaskonfiguration gehen diese Elemente so gut wie keine Reaktionen ein (einige wenige Xenon-Fluor-Verbindungen u.a. gibt es).

Das Neonatom besitzt 10 Elektronen: zwei auf der ersten und acht Valenzelektronen auf der zweiten/äußersten Schale. Ausnahme: Das Heliumatom, welches nur eine Schale besitzt, erreicht den Edelgaszustand bereits mit zwei Elektronen (Elektronenduet).

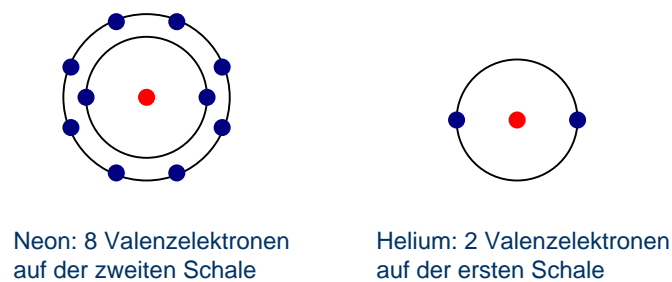


Abb. 1.8: Edelgase

Als Inertgase [lat. inert: untätig, träge] finden unter anderem Stickstoff (z.B. als Spülgas bei Ofenprozessen) und Argon (z.B. bei Sputterprozessen) in der Halbleiterherstellung Anwendung.

1.5 Leiter - Nichtleiter - Halbleiter

1.5.1 Leiter

Leiter sind generell Stoffe, die die Eigenschaft haben verschiedene Energiearten weiterzuleiten. Im Folgenden wird dabei die Leitfähigkeit des elektrischen Stroms betrachtet.

Metalle:

Die Leitfähigkeit von Metallen beruht auf den freien Elektronen die bei der Metallbindung als Elektronengas vorliegen. Bereits mit wenig Energie werden genug Elektronen von den Atomen gelöst um eine Leitfähigkeit zu erreichen.

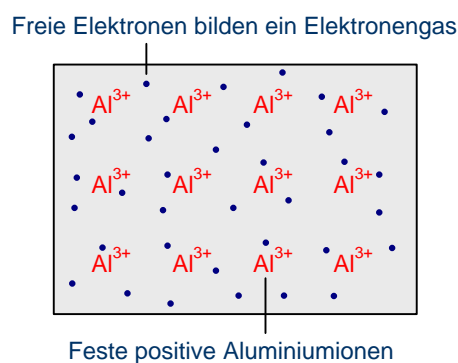


Abb. 1.9: Metallbindung: Feste Atomrümpfe und freie Valenzelektronen

Die Leitfähigkeit hängt unter anderem von der Temperatur ab. Steigt diese an, schwingen die Atomrümpfe immer stärker, so dass die Elektronen in ihren Bewegungen behindert werden. Folge, der Widerstand steigt. Die besten Leiter, Gold und Silber, werden auf Grund der hohen Kosten relativ selten eingesetzt (Gold u.a. bei der Kontaktierung der fertigen Chips). Die Alternativen in der Halbleitertechnologie zur Verdrahtung der einzelnen Komponenten eines Chips sind Aluminium und Kupfer.

Salze:

Neben Metallen können auch Salze elektrischen Strom leiten. Freie Elektronen gibt es hier jedoch nicht. Die Leitfähigkeit beruht auf den Ionen die sich beim Schmelzen oder Lösen von Salzen aus dem Gitterverbund lösen und frei beweglich sind.

1.5.2 Nichtleiter

Nichtleiter besitzen keine freien Ladungsträger in Form von Elektronen oder Ionen. Nichtleiter nennt man auch Isolatoren.

Atombindung:

Die Atombindung beruht auf gemeinsamen Elektronenpaaren von Nichtmetallen. Die Elemente mit Nichtmetallcharakter haben alle das Bestreben Elektronen aufzunehmen, somit sind keine freien Elektronen vorhanden die eine Leitfähigkeit bewirken könnten.

Ionenbindung:

Im festen Zustand sind Ionen in einem Gitterverbund angeordnet. Durch elektrische Kräfte werden die Teilchen zusammengehalten. Es sind keine freien Ladungsträger für den Stromfluss vorhanden. So können Stoffe, die sich aus Ionen zusammensetzen, sowohl Leiter als auch Nichtleiter sein.

1.5.3 Halbleiter

Halbleiter sind Feststoffe, deren Leitfähigkeit zwischen der von Leitern und Nichtleitern liegt. Durch Elektronenaustausch gleichartiger Atome, um das Elektronenoktett zu vervollständigen, ordnen sich diese als Gitterstruktur an. Im Gegensatz zu Metallen nimmt die Leitfähigkeit mit steigender Temperatur - bis zu einem gewissen Maß - zu.

Durch den Temperaturanstieg brechen Bindungen auf und Elektronen werden freige-

setzt. An der Stelle an der sich das Elektron befand verbleibt ein so genanntes Defektelektron (auch Loch).

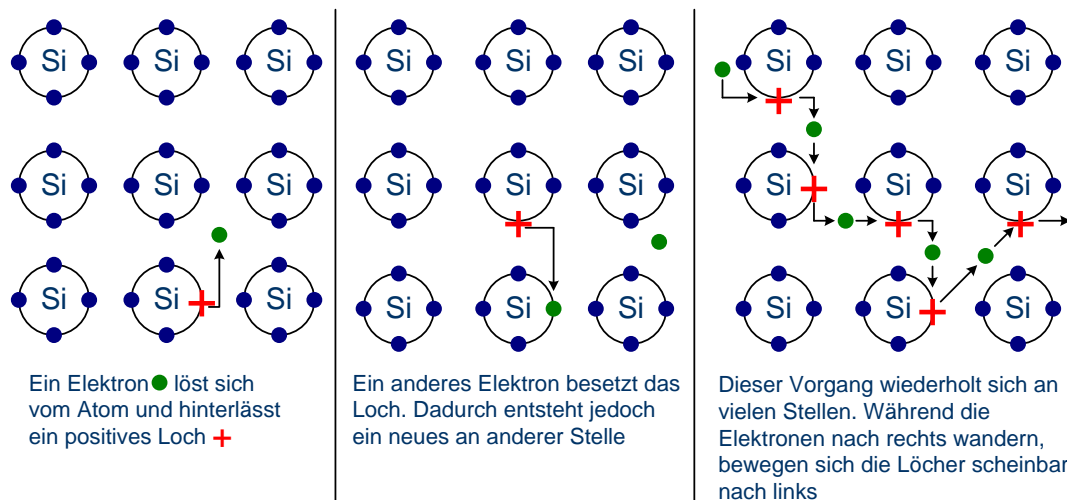


Abb. 1.10: Ausschnitt aus einem Siliciumkristall

Der Elektronenfluss beruht auf der Eigenleitfähigkeit von Halbleitern. Am so genannten Bändermodell erkennt man, warum sich Halbleiter so verhalten.

1.5.4 Das Bändermodell

Das Bändermodell ist ein Energieschema, mit Hilfe dessen man die Leitfähigkeit von Leitern, Isolatoren und Halbleitern beschreiben kann. Das Modell besteht aus zwei Energiebändern (Valenz- und Leitungsband) und der Bandlücke. Die Valenzelektronen - die als Ladungsträger dienen - befinden sich im Valenzband; das Leitungsband ist im Grundzustand nicht mit Elektronen besetzt. Zwischen den beiden Energiebändern befindet sich die Bandlücke, ihre Breite beeinflusst u.a. die Leitfähigkeit von Stoffen.

Die Energiebänder

Betrachtet man ein einzelnes Atom, so gibt es nach dem Bohrschen Atommodell scharf voneinander getrennte Energieniveaus, die von Elektronen besetzt werden können. Befinden sich mehrere Atome nebeneinander, so stehen diese miteinander in Wechselwirkung und die diskreten Energieniveaus werden aufgefächert, sie verschieben sich nach oben und unten. In einem Siliciumkristall gibt es ca. 10^{23} Atome pro Kubikzentimeter, so dass die einzelnen Energieniveaus nicht mehr von einander unterscheidbar sind und breite Energiebereiche bilden.

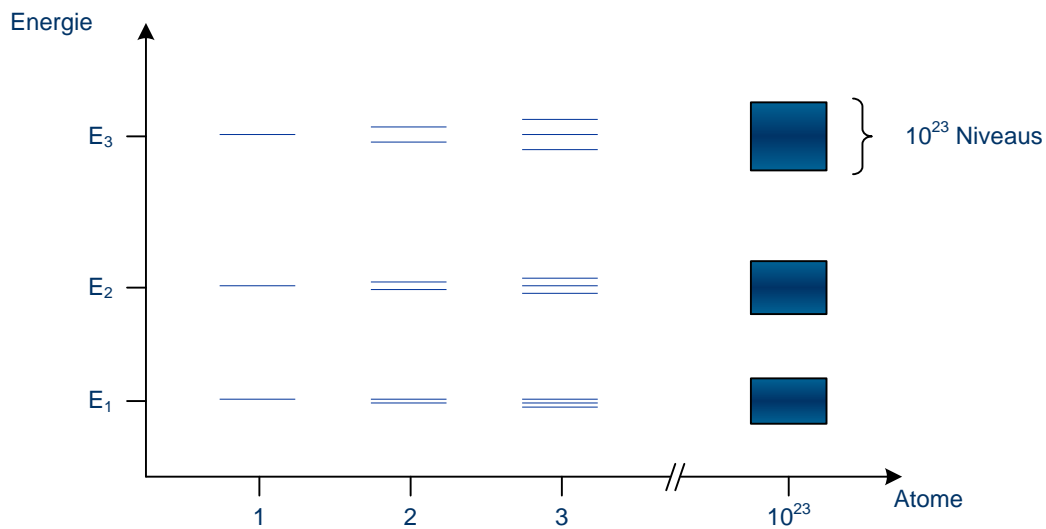


Abb. 1.11: Aufspaltung von Energieniveaus durch Wechselwirkung von Atomen

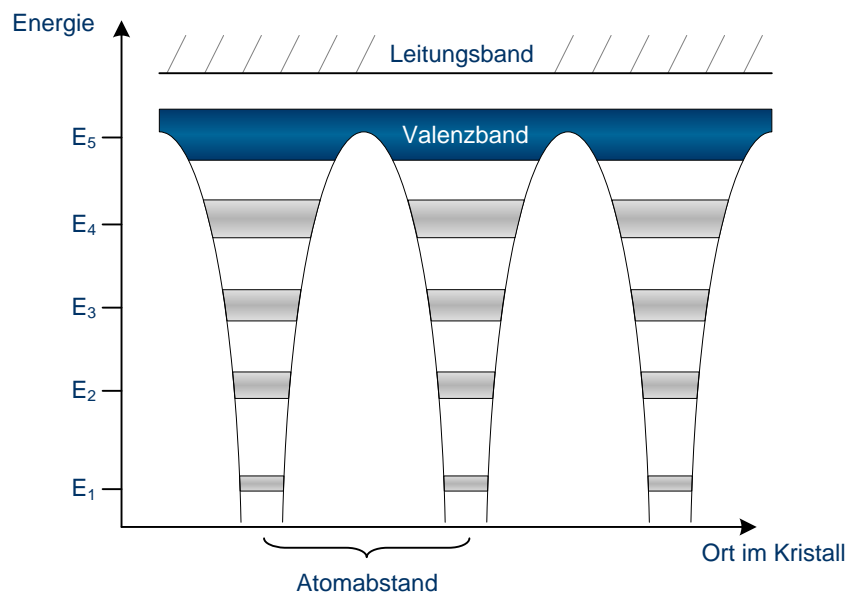


Abb. 1.12: Energiebänder durch in Wechselwirkung stehende Atome

Die Breite der Energiebänder hängt davon ab, wie stark die Elektronen an das Atom gebunden sind. Die Valenzelektronen im höchsten Energieniveau wechselwirken stark mit denen der Nachbaratome und können relativ leicht vom Atom gelöst werden, bei einer sehr großen Anzahl an Atomen lässt sich ein einzelnes Elektron nicht mehr einem bestimmten Atom zuordnen. In Folge dessen verschmelzen die Energiebänder

der einzelnen Atome zu einem kontinuierlichen Band, dem Valenzband.

Das Bändermodell bei Leitern:

Bei Leitern ist das Valenzband entweder nicht voll mit Elektronen besetzt, oder das gefüllte Valenzband überlappt sich mit dem leeren Leitungsband. In der Regel treffen beide Zustände gleichzeitig zu, die Elektronen können sich also im nur teilweise besetzten Valenzband oder in den zwei sich überlappenden Bändern bewegen. Die Bandlücke, die sich zwischen Valenz- und Leitungsband befindet, existiert dann nicht.

Das Bändermodell bei Nichtleitern:

Bei Isolatoren ist das Valenzband durch die Bindungen der Atome voll mit Elektronen besetzt. Sie können sich darin nicht bewegen, da sie zwischen den Atomen „eingesperrt“ sind. Um leiten zu können müssten sich die Elektronen aus dem voll besetzten Valenzband in das Leitungsband bewegen. Das verhindert die Bandlücke, die zwischen Valenz- und Leitungsband liegt.

Nur mit sehr großem Energieaufwand (falls überhaupt möglich) kann diese Lücke überwunden werden (in der Bandlücke darf sich nach den Gesetzen der Quantenphysik kein Elektron aufhalten).

Das Bändermodell bei Halbleitern:

Auch bei Halbleitern gibt es diese Bandlücke, diese ist im Vergleich zu Isolatoren aber so klein, dass bereits bei Raumtemperatur Elektronen aus dem Valenzband in das Leitungsband gelangen. Die Elektronen können sich hier nun frei bewegen und stehen als Ladungsträger zur Verfügung. Jedes Elektron hinterlässt außerdem ein Loch im Valenzband, welches von anderen Elektronen im Valenzband besetzt werden kann. Somit erhält man wandernde Löcher im Valenzband, die als positive Ladungsträger angesehen werden können.

Es treten immer Elektronen-Loch-Paare auf, es gibt also ebenso viele negative wie positive Ladungen, der Halbleiterkristall ist insgesamt neutral. Ein reiner, undotierter Halbleiter wird als intrinsischer Halbleiter bezeichnet, pro Kubikzentimeter gibt es in etwa 10^{10} freie Elektronen und Löcher (bei Raumtemperatur).

Da die Elektronen immer den energetisch günstigsten Zustand annehmen, fallen sie ohne Energiezufuhr wieder in das Valenzband zurück und rekombinieren mit den Löchern. Bei einer bestimmten Temperatur stellt sich ein Gleichgewicht zwischen den ins Leitungsband gehobenen und den zurückfallenden Elektronen ein. Mit zunehmender Temperatur erhöht sich die Anzahl der Elektronen, die die Bandlücke überspringen

können. Mit steigender Temperatur nimmt also die Leitfähigkeit von Halbleitern zu.

Da die Breite der Bandlücke einer bestimmten Energie und somit einer bestimmten Wellenlänge entspricht, versucht man, die Bandlücke gezielt zu verändern um so bestimmte Farben bei Leuchtdioden zu erhalten. Dies kann u.a. durch Kombination verschiedener Stoffe erreicht werden. Galliumarsenid (GaAs) hat eine Bandlücke von 1,4 eV (Elektronenvolt, bei Raumtemperatur) und strahlt somit rotes Licht ab.

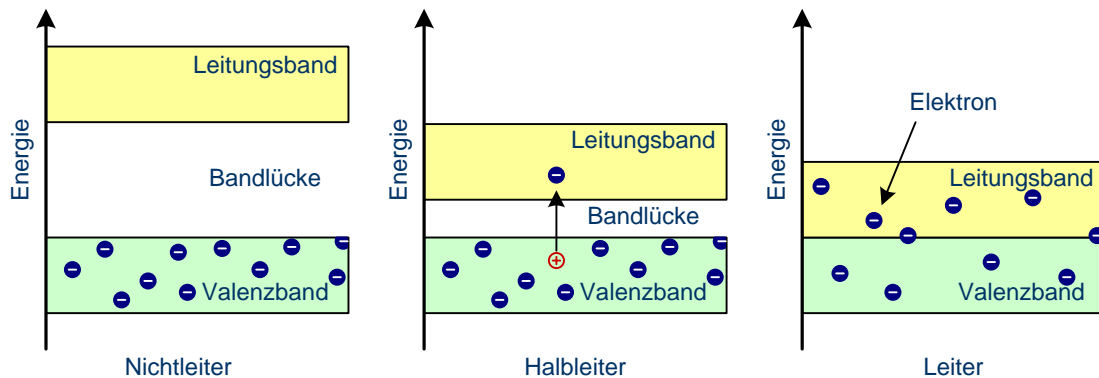


Abb. 1.13: Das Bändermodell

Die Eigenleitfähigkeit von Halbleiter ist für die Funktionsweise von Bauelementen uninteressant, da sie sehr stark von der zugeführten Energie abhängt. Sie ändert sich also auch mit der Betriebstemperatur, eine mit Metallen vergleichbare Leitfähigkeit stellt sich zudem erst mit sehr hohen Temperaturen ein (mehrere Hundert Grad Celsius). Um die Leitfähigkeit von Halbleitern gezielt zu beeinflussen, können Fremdatome in das regelmäßige Siliciumgitter eingebaut, und damit die Ladungsträgerkonzentration von Elektronen und Löchern eingestellt werden.

1.6 Dotieren: n- und p-Halbleiter

1.6.1 Dotieren

Dotieren bedeutet das Einbringen von Fremdatomen in einen Halbleiterkristall zur gezielten Veränderung der Leitfähigkeit. Zwei der wichtigsten Stoffe, mit denen Silicium dotiert werden kann, sind Bor (3 Valenzelektronen = 3-wertig) und Phosphor (5 Valenzelektronen = 5-wertig). Weitere sind: Aluminium, Indium (3-wertig) und Arsen, Antimon (5-wertig).

Die Dotierelemente werden in die Gitterstruktur des Halbleiterkristalls eingebaut, die Anzahl der Außenelektronen bestimmt die Art der Dotierung. Elemente mit 3 Valenzelektronen werden zur p-Dotierung benutzt, 5-wertige Elemente zur n-Dotierung. Die Leitfähigkeit eines gezielt verunreinigten Siliciumkristalls kann so um den Faktor 10^6 erhöht werden.

1.6.2 n-Dotierung

Das 5-wertige Dotierelement besitzt ein Außenelektron mehr als die Siliciumatome. Vier Außenelektronen verbinden sich mit je einem Siliciumatom, das fünfte ist frei beweglich und dient als Ladungsträger. Dieses freie Elektron benötigt sehr viel weniger Energie um vom Valenzband in das Leitungsband gehoben zu werden, als die Elektronen, die die Eigenleitfähigkeit des Siliciums verursachen. Das Dotierelement, welches ein Elektron abgibt, wird als Elektronendonator (donare, lat. = geben) bezeichnet.

Die Dotierelemente werden durch die Abgabe negativer Ladungsträger positiv geladen und sind fest im Gitter eingebaut, es bewegen sich nur die Elektronen. Dotierte Halbleiter, deren Leitfähigkeit auf freien (negativen) Elektronen beruht sind n-leitend bzw. n-dotiert. Auf Grund der höheren Anzahl der freien Elektronen werden diese auch als Majoritätsladungsträger bezeichnet, freie bewegliche Löcher dagegen als Minoritätsladungsträger.

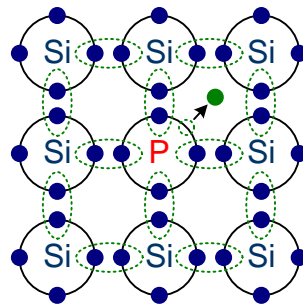


Abb. 1.14: Das Phosphor-Atom gibt sein 5. Elektron ab, das keine Bindung eingehen kann. Es dient als freier Ladungsträger.

Arsen wird als Alternative zu Phosphor verwendet, da dessen Diffusionskoeffizient geringer ist. Das bedeutet, dass der Dotierstoff bei späteren Prozessschritten weniger stark diffundiert und die Dotierung somit an der Stelle bleibt, wo sie eingebracht wurde.

1.6.3 p-Dotierung

Im Gegensatz zum freien Elektron bei der Dotierung mit Phosphor bewirken 3-wertige Dotierelemente genau das Gegenteil. Sie können ein zusätzliches Außenelektron aufnehmen und hinterlassen so ein Loch im Valenzband der Siliciumatome. Dadurch werden die Elektronen im Valenzband beweglich. Die Löcher bewegen sich in entgegengesetzter Richtung zur Elektronenbewegung. Die dazu nötige Energie beträgt bei Indium als Dotierelement nur 1 % der Energie die nötig ist, um die Valenzelektronen der Siliciumatome in das Leitungsband zu heben.

Durch die Aufnahme eines Elektrons wird das Dotierelement einfach negativ geladen; solche Dotieratome nennt man Elektronenakzeptor (acceptare, lat. = aufnehmen). Auch hier ist das Dotierelement fest im Kristallgitter eingebaut, es bewegt sich nur die positive Ladung. P-leitend oder p-dotiert nennt man diese Halbleiter weil die Leitfähigkeit auf **p**ositiven Löchern beruht. Analog zu n-dotierten Halbleitern, sind hier die Löcher die Majoritätsladungsträger, freie Elektronen die Minoritätsladungsträger.

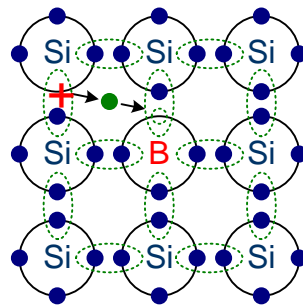


Abb. 1.15: Die freie Stelle am Bor-Atom wird von anderen Elektronen aufgefüllt. Dabei entstehen an anderer Stelle neue Löcher. Diese Leerstellen wandern scheinbar entgegengesetzt zum Elektronenfluss

Nach außen sind dotierte Halbleiter elektrisch neutral. Die Bezeichnungen n- bzw. p-Dotierung beziehen sich nur auf die Majoritätsladungsträger.

N- und p-dotierte Halbleiter verhalten sich in Bezug auf den Stromfluss annähernd gleich. Mit steigender Anzahl an Dotierelementen nimmt auch die Zahl von Ladungsträgern im Halbleiterkristall zu. Dabei genügt schon eine sehr geringe Menge an Dotierelementen. Schwach dotierte Siliciumkristalle enthalten nur 1 Fremdatom pro 10^9 Siliciumatome, bei den höchsten Dotierungen ist das Verhältnis von Fremdatomen zu Siliciumatomen z.B. 1 zu 1000.

1.6.4 Bänderschema bei dotierten Halbleitern

Bei n-dotierten Halbleitern steht durch das Einbringen eines Dotierelements mit fünf Außenelektronen ein Elektron im Kristall zur Verfügung das nicht gebunden ist, und so mit vergleichsweise geringer Energie in das Leitungsband gehoben werden kann. Somit findet man in n-dotierten Halbleitern ein Donatorniveau nahe der Leitungsbandkante, die zu überwindende Bandlücke ist sehr klein.

Durch das Einbringen eines 3-wertigen Dotierelements steht bei p-dotierten Halbleitern eine Leerstelle zur Verfügung, die schon mit geringer Energie von einem Elektron aus dem Valenzband besetzt werden kann. Bei p-dotierten Halbleitern befindet sich somit ein Akzeptorniveau nahe der Valenzbandkante.

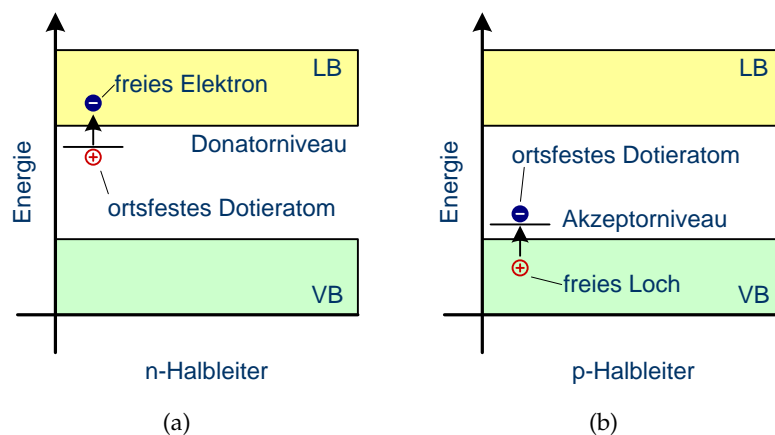


Abb. 1.16: Bändermodell von n- und p-Dotierten Halbleitern

1.7 Der p-n-Übergang

1.7.1 p-n-Übergang ohne angelegte Spannung

Dioden sind elektrische Bauteile, die durch Anlegen von Spannung den Strom in der einen Richtung leiten (Durchlassrichtung) und in der anderen sperren (Sperrrichtung). Mit n- und p-dotierten Halbleitern kann man dieses Verhalten erreichen.

Der p-n-Übergang ist der Übergangsbereich aneinander liegender n- und p-dotierter Halbleiterkristalle. In diesem Bereich gibt es keine freien Ladungsträger, da die freien

Elektronen des n-Leiters und die freien Löcher des p-dotierten Kristalls in der Nähe der Kontaktfläche der zwei Kristalle miteinander rekombinieren, d.h. die Elektronen besetzen die freien Löcher. Diese Ladungsträgerbewegung (Diffusion) ergibt sich in Folge eines Konzentrationsgefälles. Da es im p-Gebiet nur wenige Elektronen und im n-Gebiet nur wenige Löcher gibt, wandern die Majoritätsladungsträger in den jeweils andersartig dotierten Halbleiterkristall.

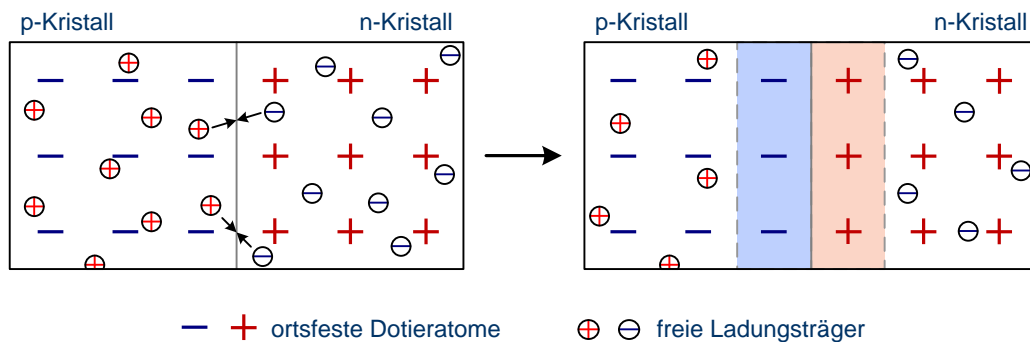


Abb. 1.17: p-n-Übergang ohne angelegte Spannung

Die Gebiete in Nähe der Grenzschicht sind auf Grund der abgewanderten freien Ladungsträger positiv (n-Kristall) bzw. negativ (p-Kristall) geladen. Je mehr Ladungsträger rekombinieren, desto größer wird diese Verarmungs- oder Raumladungszone (RLZ) und damit die Spannungsdifferenz von n- zu p-Kristall. Bei einer bestimmten Höhe dieses Potentialgefälles kommt die Rekombination der Löcher und Elektronen zum Erliegen, die Ladungsträger können das elektrische Feld nicht mehr überwinden. Bei Silicium liegt diese Grenze bei etwa 0,7 V.

1.7.2 p-n-Übergang mit angelegter Spannung

Wird am n-Kristall eine positive und am p-Kristall eine negative Spannung angelegt, so zeigen das elektrische Feld im Inneren und das durch die Spannungsquelle erzeugte in die gleiche Richtung. Das Feld am p-n-Übergang wird damit verstärkt. Die jeweils entgegengesetzt geladenen freien Ladungsträger werden von den Polen der Spannungsquelle angezogen, dadurch wird die Sperrschicht vergrößert und es ist kein Stromfluss möglich.

Polt man die angelegte Spannung an den Halbleiterkristallen um, überlagert das durch die Spannungsquelle erzeugte elektrische Feld das innere in entgegengesetzter Richtung und schwächt es ab. Wird das innere Feld vollständig vom äußeren abgebaut, fließt

ßen ständig neue Ladungsträger von der Stromquelle zur Sperrschicht und können hier fortlaufend rekombinieren: es fließt Strom.

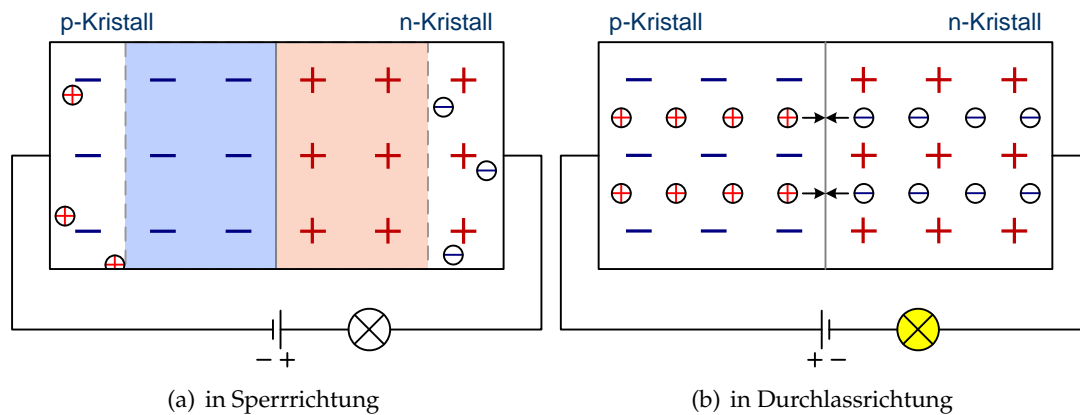


Abb. 1.18: p-n-Übergang mit angeschlossener Spannungsquelle

Die Diode lässt sich auf Grund dieses Verhaltens als Gleichrichter verwenden: zur Umwandlung von Wechselstrom in Gleichstrom. Bereiche, in denen p- und n-dotierte Halbleiterkristalle in Kontakt stehen, kommen in vielen elektrischen Bauelementen der Halbleitertechnologie vor.

1.8 Aufbau eines n-Kanal-FET

1.8.1 Allgemeiner Aufbau

Ein Transistor ist ein elektronisches Halbleiterbauelement das zum Schalten oder Verstärken von Strom verwendet werden kann. Der Strom kann über zwei Anschlüsse fließen (Drain, Source), während der dritte (Gate) zur Steuerung dient. Neben dem Feldeffekttransistor (FET), der hier beschrieben wird, gibt es noch einen weiteren grundlegenden Transistortyp, den Bipolartransistor. Bei ihm heißen die Anschlüsse Emitter (Source beim FET), Basis (Gate) und Kollektor (Drain). Die Funktionsweise des Bipolartransistors beruht auf Ladungsträgern beider Polaritäten (daher bipolar), Löchern und Elektronen. Beim Feldeffekttransistor, auch als unipolarer Transistor bezeichnet, sind abhängig von der Bauart entweder Elektronen *oder* Löcher am Stromtransport beteiligt.

Bei dem nachfolgend beschriebenen Transistor handelt es sich um einen sogenannten

MOSFET (engl. metal oxide semiconductor field-effect transistor, Metall-Oxid-Halbleiter-FET). Obwohl heute meist hochdotiertes Polysilicium als Gatematerial Verwendung findet und kein Aluminium mehr zum Einsatz kommt, wird auch bei diesem Transistortyp nach wie vor die Bezeichnung MOSFET benutzt. Besser wäre in diesem Fall die Bezeichnung IGFET (engl. insulated gate FET, FET mit isoliertem Gate). Bei neuartigen Transistoren mit High-k-Metal-Gate-Technologie ist die Bezeichnung MOSFET dagegen wieder korrekt, sofern als Isolator weiterhin ein Oxid verwendet wird.

Der Transistor ist das grundlegenden Bauelemente in der Halbleiterfertigung, in modernen Mikrochips finden sich mehrere hundert Millionen bis Milliarden Transistoren. Durch die Kombination mehrerer Transistoren können sämtliche logische Gatter realisiert werden, um aus Eingangssignalen entsprechende logische Ausgangssignale zu erhalten. Dadurch bilden Transistoren das Herzstück eines jeden Mikroprozessors, Speicherchips usw. Der Transistor ist die von der Menschheit in der höchsten Gesamtstückzahl produzierte technische Funktionseinheit, und aus dem heutigen Leben nicht mehr wegzudenken.

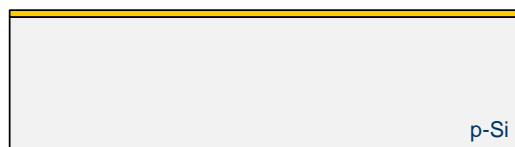
Der Transistor wird in der Produktion Schicht für Schicht aufgebaut. Dabei steht hier der grundlegende Aufbau eines einfachen MOSFETs im Vordergrund, die verschiedenen Möglichkeiten zur Realisierung dieser Schichten folgen in den späteren Kapiteln.

1.8.2 Aufbau eines n-Kanal-FET

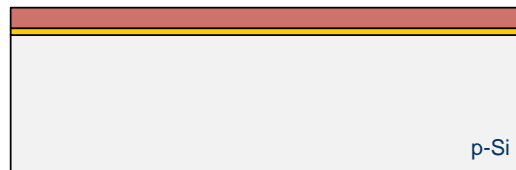
1. Substrat: Grundlage für einen n-Kanal-Feldeffekttransistor ist ein p-dotiertes Siliciumsubstrat, als Dotierstoff dient Bor.



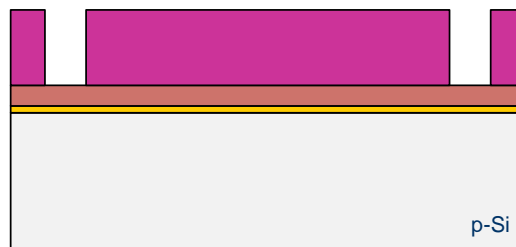
2. Oxidation: Auf dem Substrat wird Siliciumdioxid SiO_2 (das Gateoxid, kurz GOX, z.B. 40 nm) in einer Trockenoxidation erzeugt. Es dient zur Isolation zwischen dem Gate und dem Substrat.



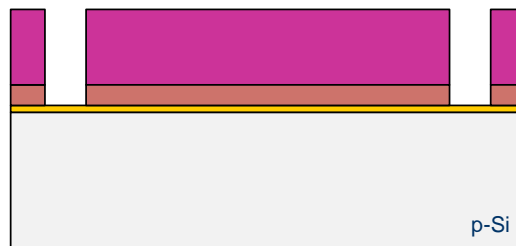
3. Abscheidung: In einem LPCVD-Prozess wird Nitrid abgeschieden, es dient später bei der Feldoxidation als Maskierung.



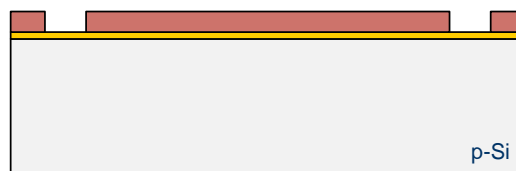
4. Fototechnik: Auf dem Nitrid wird ein Fotolack aufgebracht, belichtet und entwickelt. So erhält man eine strukturierte Lackschicht, die als Ätzmaske dient.



5. Ätzen: Der Lack maskiert das Nitrid, lackfreie Bereiche werden mittels reaktivem Ionenätzen entfernt.

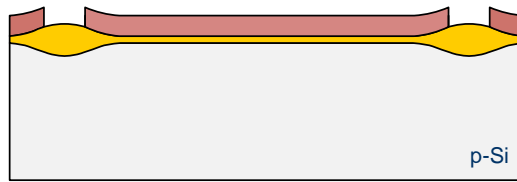


6. Lackentfernen: Nach der Strukturübertragung der Lackmaske in das Nitrid, wird der Resist mit einer Entwicklerlösung nasschemisch entfernt.



7. Oxidation: Nur auf den Nitrid freien Bereichen findet eine thermische Nassoxidation statt. Das Feldoxid (FOX, z.B. 700 nm) dient zur seitlichen Isolation zu benachbarten

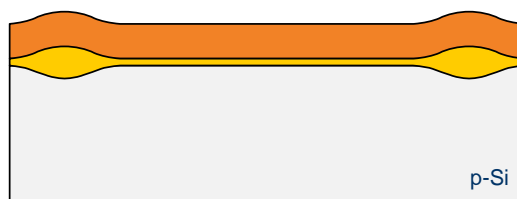
Bauteilen.



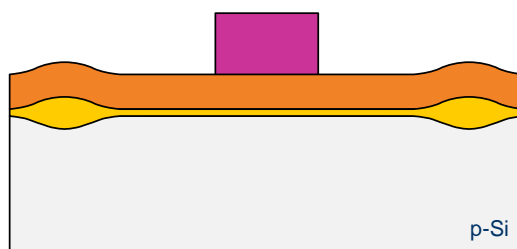
8. Ätzen: Nach der Oxidation wird das Nitrid in einem nasschemischen Ätzprozess entfernt.



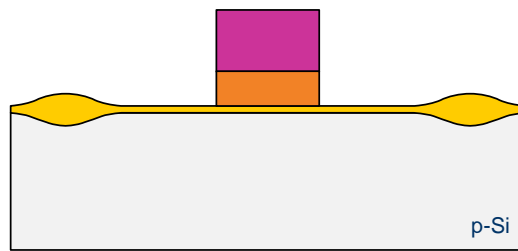
9. Abscheidung: In einem LPCVD-Prozess wird polykristallines Silicium abgeschieden.



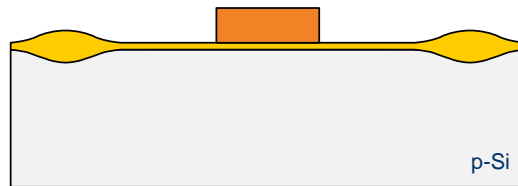
10. Fototechnik: Auf dem Polysilicium wird eine Lackschicht als Ätzmaske strukturiert.



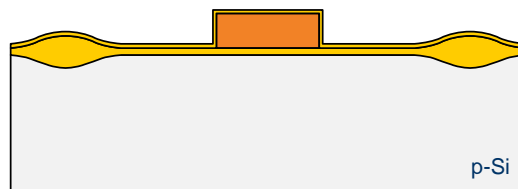
11. Ätzen: Der Fotolack dient wiederrum als Lackmaske, in einem reaktiven Ionenätzschritt wird das Silicium strukturiert. Es dient als Gateelektrode (z.B. 300 nm) zur Steuerung des Transistors.



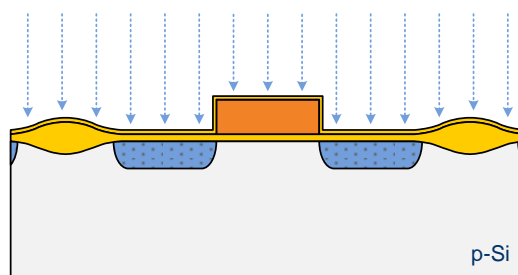
12. Lackentfernen: Der Lack wird nach dem Ätzschritt wieder nasschemisch entfernt.



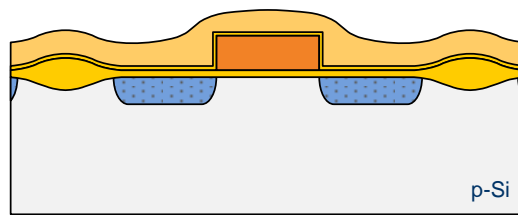
13. Oxidation: Nachfolgend wird ein dünnes Oxid, das Postoxid abgeschieden.



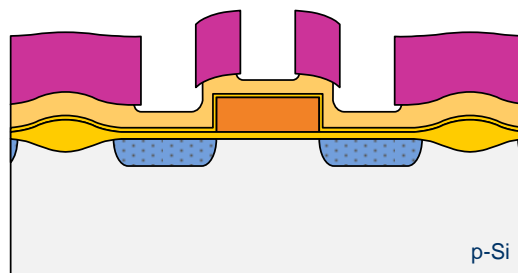
14. Ionenimplantation: In einem Implantationsschritt mit Phosphorionen werden das Source- und Draingebiet n-dotiert. Da die Gateelektrode als Implantationsmaske dient und so die Weite des n-Kanals zwischen Source und Drain vorgegeben ist, bezeichnet man dies als Selbstjustierung.



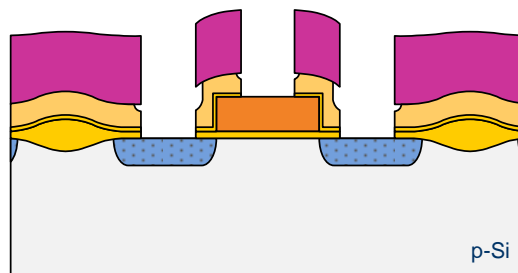
15. Oxidation: Als Isolation zu darüberliegenden Metallisierungsschichten, wird ein Oxid (das Zwischenoxid, kurz ZOx, z.B. 700 nm) abgeschieden. Dies geschieht in einem LPCVD-Prozess mit TEOS, welches eine gute Kantenbedeckung bietet.



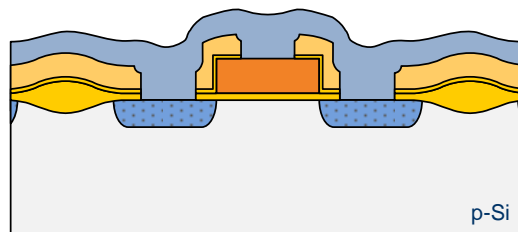
16. Fototechnik und Ätzen: Darüber wird eine weitere Lackschicht strukturiert und in einem isotropen Ätzprozess die Kanten der Kontaktlöcher verrundet.



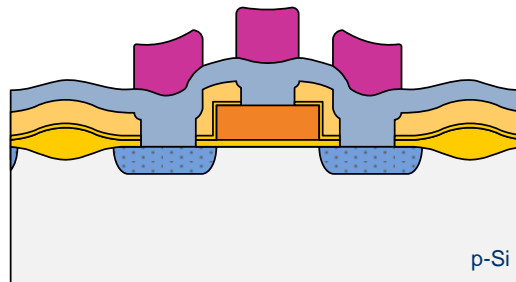
17. Ätzen: Anschließend werden die Kontaktlöcher in einem anisotropen Ätzprozess bis zu den n-dotierten Gebieten freigelegt.



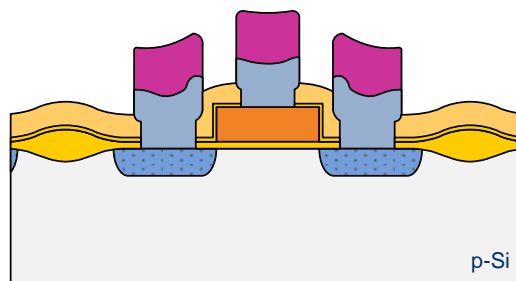
18. Metallisierung: Die Kontaktlöcher werden in einem Sputterprozess mit Aluminium aufgefüllt.



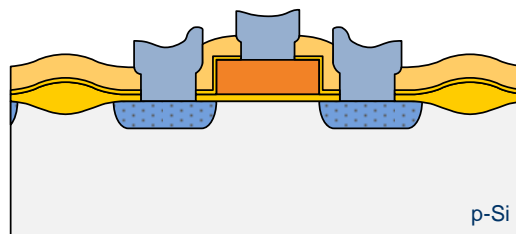
19. Fototechnik: In einem letzten Lithografieschritt wird eine neue Lackschicht strukturiert.



20. Ätzen: Die Lackstrukturen werden in einem anisotropen Trockenätzschritt in die darunterliegende Metallisierungsebene übertragen.



21. Lackentfernen: Abschließend wird der Resist entfernt, zurück bleiben Aluminiumleiterbahnen.



1.8.3 Funktionsweise

Anreicherungs-FET:

Ohne eine positive Spannung am Gate stehen zwischen Source und Drain keine freien Ladungsträger in Form von Elektronen Verfügung, da das Substrat p-dotiert ist. Im

stationären Zustand sind hier also Löcher die Majoritätsladungsträger und die Elektronen in der Minderheit.

Durch eine positive Spannung am Gate werden Elektronen durch das elektrische Feld aus dem Substrat angezogen (Löcher dementsprechend verdrängt) und bilden so einen leitenden n-Kanal zwischen Source und Drain. Die isolierende Siliciumdioxidschicht verhindert einen Stromfluss zwischen Substrat und Gate. Da der Transistor den Stromfluss ohne angelegte Spannung sperrt nennt man den Transistor auch selbstsperrend.

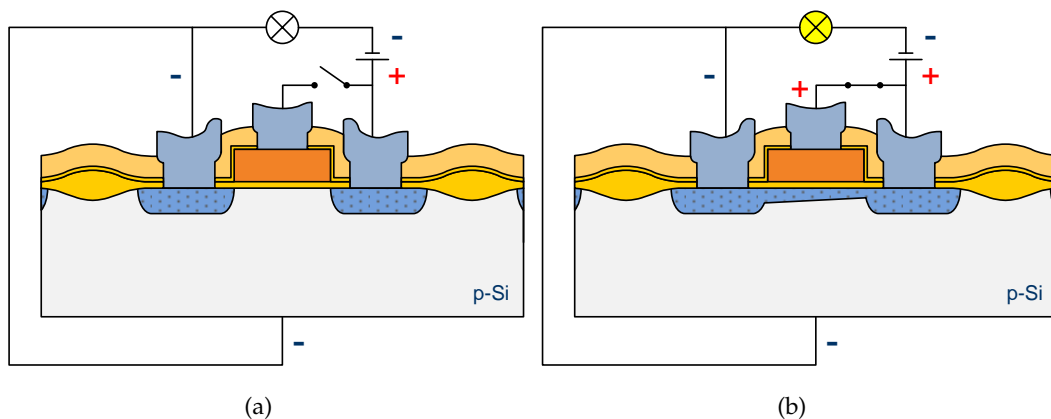


Abb. 1.19: Anreicherungs-Feldeffekttransistor

Verarmungs-FET:

Durch eine leichte n-Dotierung zwischen Source und Gate erreicht man, dass ein Transistor auch ohne Gatespannung leitend ist (Spannung zwischen Source und Drain reicht aus). So genannte Verarmungs-FETs, oder auch selbstleitende Transistoren, sperren nur dann, wenn am Gate eine negativere Spannung als am Source-Anschluss anliegt. Dadurch werden die Elektronen, die sich unter dem Gate befinden, verdrängt - die leitende Elektronenbrücke geht verloren.

1.9 Aufbau eines Bipolartransistors

1.9.1 Allgemeiner Aufbau

Der zweite wichtige Transistortyp neben dem Feldeffekttransistor ist der Bipolartransistor. Seine Funktionsweise beruht auf beiden Ladungsträgern (bipolar), Elektronen

und Löffern. Bipolartransistoren sind schneller als Feldeffekttransistoren, beanspruchen jedoch mehr Platz und sind somit teurer.

Bipolartransistoren bestehen im Wesentlichen aus zwei gegeneinander geschalteten p-n-Übergängen mit der Schichtfolge n-p-n oder p-n-p. Die Anschlüsse des Bipolartransistors werden als Emitter (E), Basis (B) und Kollektor (C) bezeichnet, Emitter und Kollektor besitzen jeweils die gleiche Dotierungsart. Zwischen den beiden Anschlüssen befindet sich die sehr dünne Basisschicht, die dementsprechend jeweils anders dotiert ist.

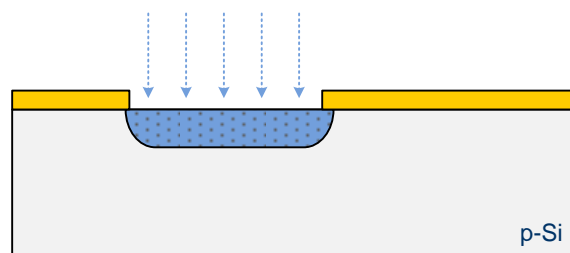
Beschrieben wird hier ein npn-Transistor in Standard Buried Collector-Bauweise (SBC, vergrabener Kollektor), die Funktionsweise des pnp-Transistors ist analog dazu (die Vorzeichen der angelegten Spannung müssen lediglich vertauscht werden).

1.9.2 Aufbau eines npn-Bipolartransistors

1. Substrat: Grundlage für einen npn-Bipolartransistor ist ein p-dotiertes Siliciumsubstrat, als Dotierstoff dient Bor. Darauf wird ein dickes Oxid (z.B. 600 nm) abgeschieden.

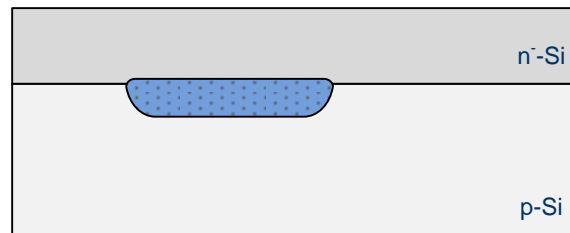


2. Buried-Layer-Implantation: Das Oxid dient als Implantationsmaske, als Dotierstoff wird Antimon Sb eingesetzt, da dies im Vergleich zu Phosphor bei späteren Diffusionsprozessen weniger stark diffundiert. Der stark n+-dotierte, vergrabene Kollektor dient als niederohmige Kontaktfläche für den Kollektoranschluss.

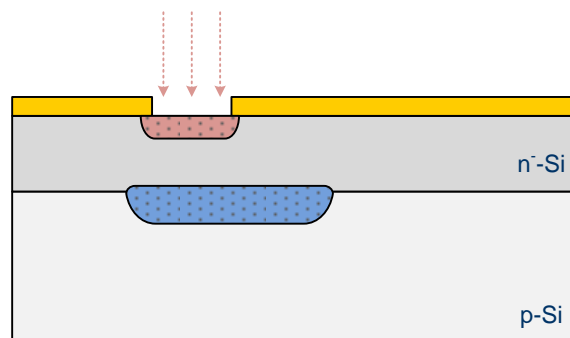


3. Homoepitaxie: In einem Epitaxieprozess wird eine hochohmige, schwach n^- -dotierte

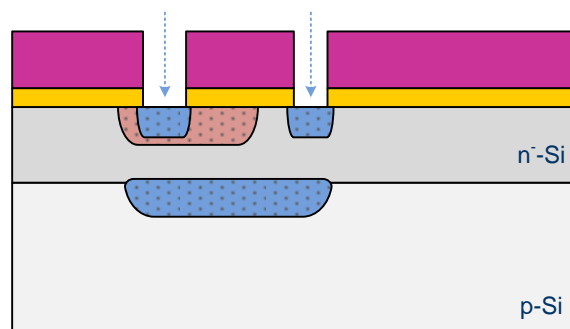
te Kollektorschicht abgeschieden (typisch $10\ \mu\text{m}$).



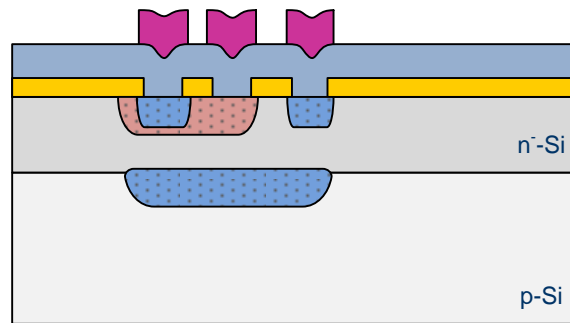
4. Basis-Implantation: Mit Borionen wird die p-dotierte Basis erzeugt, bei einem anschließenden Diffusionsschritt wird das Gebiet vergrößert.



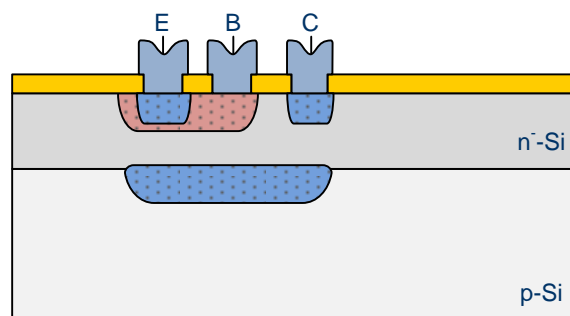
5. Emitter- und Kollektor-Implantation: Mit Phosphorionen werden die beiden stark n^+ -dotierten Emitter- und Kollektoranschlüsse eingebracht.



6. Metallisierung und Fototechnik: In einem Sputterprozess wird zur Kontaktierung der drei Anschlüsse Aluminium abgeschieden und darüber eine Lackschicht strukturiert.



7. Ätztechnik: Abschließend werden die Anschlüsse für Emitter, Basis und Kollektor in einem anisotropen Trockenätzschritt strukturiert.



1.9.3 Funktionsweise

Die beiden p-n-Übergänge werden im Folgenden als EB (Emitter-Basis) bzw. CB (Kollektor-Basis) bezeichnet. Ohne äußere Spannung bilden sich an EB und CB Raumladungszonen aus. Mit einer negativen Spannung am Emitter und einer positiven Spannung am Kollektor wird die Raumladungszone an EB abgebaut, an CB jedoch vergrößert. Wird an der Basis nun eine positive Spannung angelegt, so wird EB leitend - Elektronen gelangen in die Basisschicht. Da diese sehr dünn ist, können die Ladungsträger in den Kollektor injiziert werden, wo sie auf Grund der angelegten positiven Spannung abgesaugt werden. Somit fließt ein Strom von Emitter zu Kollektor. Nahezu alle Elektronen gelangen so bereits bei einer geringen Spannung an der Basis zum Kollektor (>95 %), was bedeutet, dass mit einem relativ kleinen Basisstrom (Emitter zu Basis) ein sehr großer Kollektorstrom (Emitter zu Kollektor) ermöglicht wird.

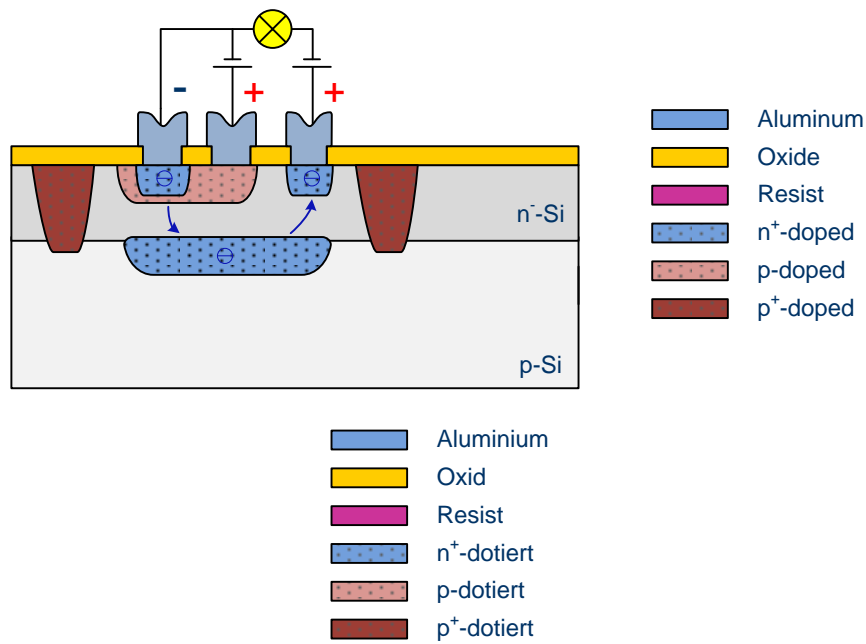


Abb. 1.20: SBC-Bipolartransistor

Die beiden tiefen p^+ -dotierten Gebiete dienen zur seitlichen Isolation gegen andere Bauteile. Neben dem Transistor ist zudem ein Widerstand notwendig (nicht in Grafik), da Bipolartransistoren nicht stromlos angesteuert werden können.

1.10 Aufbau eines FinFET

1.10.1 Allgemeiner Aufbau und Funktionsweise

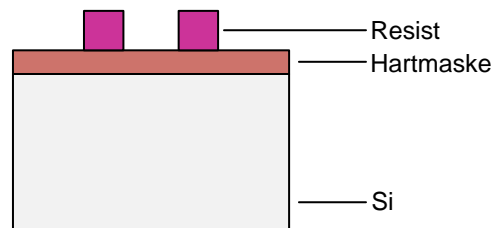
Der grundlegende Aufbau und die Funktionsweise eines FinFET unterscheiden sich nicht von einem herkömmlichen MOS-Feldeffekttransistor. So gibt es auch hier einen Source- und Drainanschluss, über die der Stromfluss erfolgt. Die Steuerung des Transistors regelt eine Gateelektrode. Im Gegensatz zum klassischen, in Planarbauweise hergestellten Feldeffekttransistor wird der Kanal zwischen Source und Drain jedoch als dreidimensionale Struktur auf dem Siliciumsubstrat erzeugt, so dass die Gateelektrode diesen von mehreren Seiten umschließen kann. So wird ein wesentlich besseres elektrisches Verhalten ermöglicht: Leckströme können reduziert und Steuerströme besser kontrolliert werden.

Die dreidimensionale Struktur erzeugt jedoch auch neue parasitäre Kapazitäten und minimale Abmessungen (engl. critical dimension), die optimiert werden müssen. Die Gatelänge wird in einem FinFET parallel zum Kanal gemessen, während die Weite des Gates der doppelten Finnenhöhe plus der Finnenbreite entspricht. Die Höhe des Kanals limitiert den Steuerstrom und die Gatekapazität, die Breite beeinflusst die Durchbruchspannung und Kurzkanaleffekte und die daraus resultierenden Größen wie den Stromverbrauch.

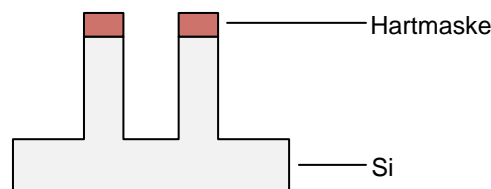
Im Folgenden wird der Aufbau eines Multigate-Transistors mit drei Gates (Tri-Gate) im Bulk-Prozess beschrieben.

1.10.2 Aufbau eines FinFET im Bulkprozess

1. Substrat: Grundlage für einen FinFET ist ein schwach p-dotiertes Siliciumsubstrat, auf dem eine Hartmaske (z.B. Siliciumnitrid) aufgebracht und eine Lackschicht strukturiert wird.

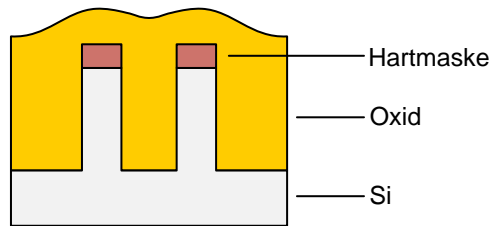


2. Ätzung der Finnen: In einem stark anisotropen Trockenätzschritt werden die freistehenden Finnen in das Siliciumsubstrat geätzt. Beim Bulk-Prozess muss die Ätzung im Gegensatz zur Herstellung auf einem SOI-Substrat mit Festzeit erfolgen, da keine Stoppsschicht im Silicium Aufschluss über die erreichte Tiefe gibt. In einem 22-nm-Prozess liegt die Breite der Finnen bei ca. 10-15 nm, die Höhe sollte idealerweise dem doppelten oder mehr entsprechen.

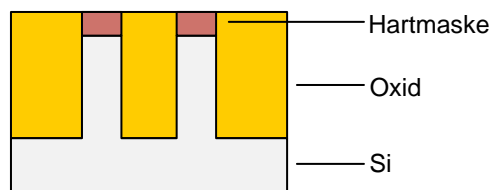


3. Oxidabscheidung: Zur Isolation erfolgt eine Oxidabscheidung, welche ein gutes

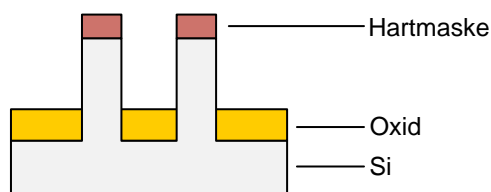
Füllverhalten von schmalen, tiefen Gräben erlauben muss.



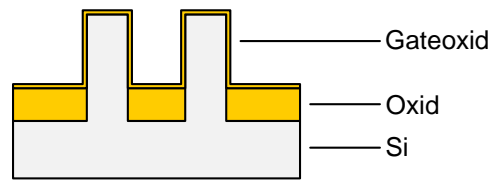
4. Planarisierung: Mittels chemisch mechanischem Polieren wird das Oxid eingeebnet. Die Hartmaske dient als Stoppschicht.



5. Rückätzung: In einem weiteren Ätzschritt mit Festzeit wird das Oxid zurückgeätzt. Zurück bleibt eine seitliche Isolation zwischen den Finnen. Da die Finnen über das Substrat in Kontakt stehen, muss ein Dotierschritt erfolgen, durch den die Siliciumstege elektrisch von einander isoliert werden (nicht dargestellt). Alternativ kann das Oxid in einem Hochtemperaturschritt in den Bodenbereich der Finnen auswachsen, so dass die Kanäle vom Siliciumsubstrat isoliert werden. Auf Grund damit verbundener komplexer Prozessschritte ist es jedoch unwahrscheinlich, dass dieser Ansatz in der Massenfertigung zum Einsatz kommen wird.



6. Gateoxid: Auf den Finnen wird das Gateoxid mittels thermischer Oxidation abgeschieden, um den Kanalbereich von der Gateelektrode zu isolieren.



7. Gateabscheidung: Abschließend wird in einem LPCVD-Prozess ein stark n^+ -dotiertes Polysilicium auf den Finnen abgeschieden. Dadurch ergeben sich mehrere Gateelektroden: je eine links und rechts vom Kanal sowie - in Abhängigkeit der Dicke des Gateoxids auf den Finnen - eine dritte darüber.

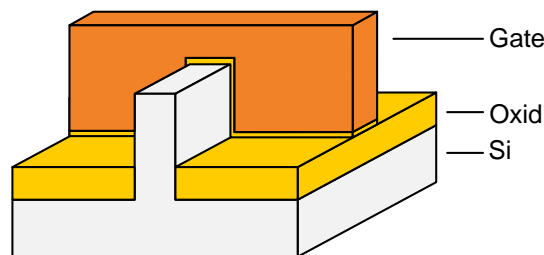


Abb. 1.21: FinFET im Bulk-Prozess

Die Einwirkung des Topgates auf den Kanal kann auch durch eine Nitridschicht zwischen Finne und Gate unterbunden werden.

Da bei einem SOI basierten Prozess bereits eine ganzflächige Oxidschicht auf dem Wafer vorhanden ist, ist die elektrische Isolation der Kanäle voneinander automatisch gegeben. Zusätzlich ist der Ätzprozess der Finnen unproblematischer, da dieser einfach auf dem Oxid gestoppt werden kann.

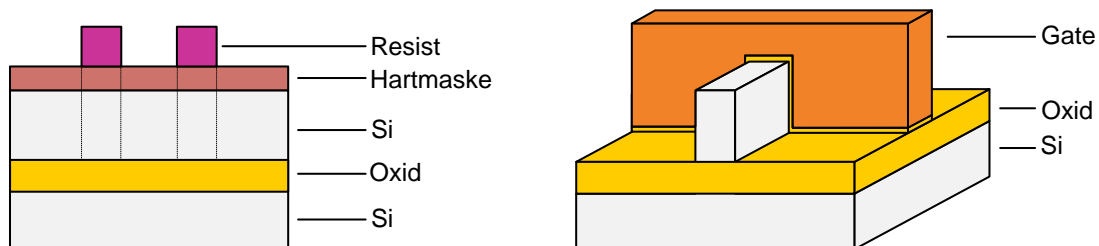


Abb. 1.22: FinFET im SOI-Prozess

2 Waferherstellung

2.1 Silicium

Silicium ist das chemische Element mit der Ordnungszahl 14 im Periodensystem der Elemente, es steht in der 4. Hauptgruppe und der 3. Periode. Silicium ist ein klassischer Halbleiter, seine Leitfähigkeit liegt also zwischen der von Leitern und Nichtleitern. Silicium (von lat. silex / silicis: Kieselstein) kommt in der Natur ausschließlich als Oxid vor: als Siliciumdioxid (SiO_2) in Form von Sand, Quarz oder als Silicat (Verbindungen von Silicium mit Sauerstoff, Metallen u.a.). Silicium gibt es also sprichwörtlich wie Sand am Meer, dementsprechend ist es ein sehr günstiges Ausgangsmaterial, dessen Wert erst mit der Verarbeitung bestimmt wird. Andere Halbleiter wie Germanium oder der Verbindungshalbleiter Galliumarsenid bieten teils deutlich bessere elektrische Eigenschaften als Silicium, die Ladungsträgerbeweglichkeit und die daraus resultierenden Schaltgeschwindigkeiten sind bei Germanium und GaAs deutlich höher. Doch Silicium hat entscheidende Vorteile gegenüber anderen Halbleitern.

Auf einem Siliciumkristall lassen sich sehr leicht Oxidschichten erzeugen, das entstandene Siliciumdioxid ist ein hochwertiger Isolator, der sich gezielt auf dem Substrat aufbringen lässt. Bei den angesprochenen Halbleitern Germanium und GaAs ist es dagegen sehr kostenintensiv ähnliche Isolationsschichten zu erzeugen. Auch die Möglichkeit, durch Dotierung ganz gezielt die Leitfähigkeit des reinen Siliciums zu verändern macht das Halbmetall so bedeutsam. Andere Stoffe sind zudem teilweise sehr giftig, und Verbindungen mit diesen Elementen nicht so langlebig und stabil wie bei Silicium. Voraussetzung für den Einsatz von Silicium in der Halbleiterfertigung ist jedoch, dass das Silicium in einer hochreinen Form als Einkristall vorliegt. Das bedeutet, dass die Siliciumatome im Kristallgitter absolut regelmäßig angeordnet sind und sich keine undefinierten Fremdatome im Kristall befinden.

Neben der einkristallinen Form gibt es noch Polysilicium (poly = viel) und amorphes Silicium (a-Si). Während das einkristalline Silicium als Wafer die Grundlage für die Mi-

kroelektronik ist, wird polykristallines Silicium in verschiedenen Bereichen als Schicht auf dem Wafer erzeugt, um bestimmte Aufgaben zu erfüllen (z.B. Maskierschicht, Gate im Transistor u.a.). Es lässt sich einfach erzeugen und leicht strukturieren. Polysilicium setzt sich aus vielen einzelnen, zu einander unregelmäßig angeordneten Einkristallen zusammen. Amorphes Silicium besitzt keinen regelmäßigen Gitteraufbau sondern eine ungeordnete Gitterstruktur und spielt in der Halbleiterfertigung keine Rolle, bietet aber Vorteile gegenüber den anderen Formen von Silicium bei der Herstellung von Dünnschichtsolarzellen.

2.2 Herstellung des Rohsiliciums

2.2.1 Herstellung von Rohsilicium

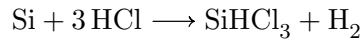
Silicium, das in der Halbleiterfertigung Verwendung findet, wird aus Quarz gewonnen. Dabei muss Sauerstoff, der sich bereits bei Raumtemperatur sehr schnell mit Silicium verbindet, und der auch beim Quarz in Verbindung mit Silicium als Siliciumdioxid vorliegt, entfernt werden. Das geschieht knapp über dem Schmelzpunkt von Silicium (1414 °C) in Öfen unter Verwendung von Kohlenstoff. Bei 1460 °C spaltet sich der Sauerstoff vom Silicium ab und reagiert mit Kohlenstoff C zu Kohlenmonoxid CO:



Zugesetztes Eisen verhindert, dass sich Silicium mit Kohlenstoff zu Siliciumkarbid verbindet. Das Kohlenmonoxid ist bei diesen Temperaturen gasförmig, und kann leicht vom flüssigen Silicium getrennt werden. Jedoch ist das Rohsilicium noch stark verunreinigt. Es enthält noch bis zu 5 % Fremdstoffe, wie z.B. Eisen, Aluminium, Phosphor und Bor. Diese Stoffe müssen in weiteren Prozessen entfernt werden.

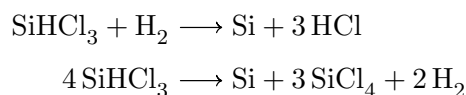
2.2.2 Reinigung von Rohsilicium

Mittels des Trichlorsilanprozesses werden viele Verunreinigungen durch Destillation herausgefiltert. Das Rohsilicium und Chlorwasserstoff HCl reagieren bei ca. 300 °C zu gasförmigem Trichlorsilan SiHCl₃ und Wasserstoff H₂:



Die Fremdstoffe, die sich ebenfalls mit dem enthaltenen Chlor verbinden, gehen erst bei höheren Temperaturen in den gasförmigen Zustand über. So lässt sich das Trichlorsilan abtrennen. Lediglich Kohlenstoff, Phosphor und Bor, die ähnliche Kondensationstemperaturen haben, können hier nicht herausgefiltert werden.

Der Trichlorsilanprozess lässt sich umkehren, so dass man anschließend das gereinigte Silicium in polykristalliner Form erhält. Dies geschieht unter Zugabe von Wasserstoff in einem Quarzgefäß, in dem sich beheizte, dünne Siliciumstäbe (Siliciumseelen) befinden, bei ca. 1100 °C:



Das Silicium schlägt sich auf den Siliciumseelen nieder, die dadurch auf über 300 mm dicke Stäbe anwachsen. Mit dem Czochralski-Verfahren könnte man dieses Polysilicium bereits in einen Einkristall umwandeln, jedoch ist der Reinheitsgrad für die Herstellung von Bauelementen noch immer nicht hoch genug.

2.2.3 Zonenreinigung

Um die Reinheit weiter zu erhöhen wird die Zonenreinigung verwendet. Um die Siliciumstäbe wird eine Spule gelegt, die hochfrequenter Wechselstrom durchfließt. Dadurch schmilzt das Silicium im Inneren der Stäbe und die Verunreinigungen sinken auf Grund der hohen Löslichkeit nach unten (die Oberflächenspannung des Siliciums verhindert, dass die Schmelze herausfließt). Durch mehrfache Wiederholung dieses Verfahrens wird der Gehalt der Fremdstoffe im Silicium so weit gesenkt, dass es zum Einkristall weiter verarbeitet werden kann.

Die Prozesse werden alle im Vakuum durchgeführt, um weitere Verunreinigungen zu verhindern.

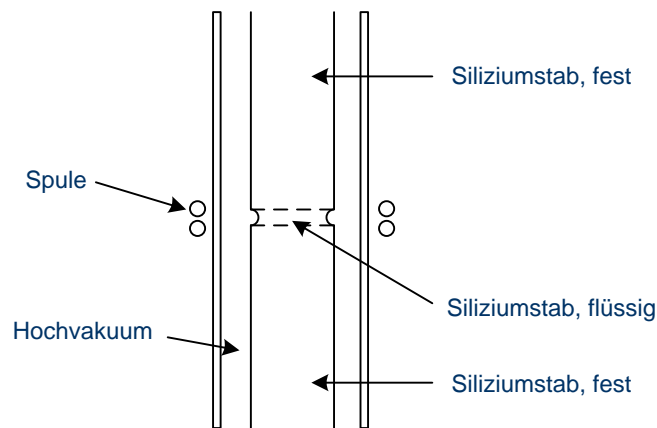


Abb. 2.1: Darstellung der Zonenreinigung

2.3 Herstellung des Einkristalls

2.3.1 Der Einkristall

Ein Einkristall (Monokristall), wie er in der Halbleiterfertigung benötigt wird, ist eine regelmäßige Anordnung von Atomen. Daneben gibt es polykristallines (Zusammensetzung von vielen kleinen Einkristallgittern) und amorphes Silicium (ungeordnete Struktur). Je nachdem, welche Lage das Gitter im Raum hat, besitzen Siliciumscheiben unterschiedliche Oberflächenstrukturen, die Einfluss auf verschiedene Parameter, wie die Ladungsträgerbeweglichkeit, von Bauelementen haben.

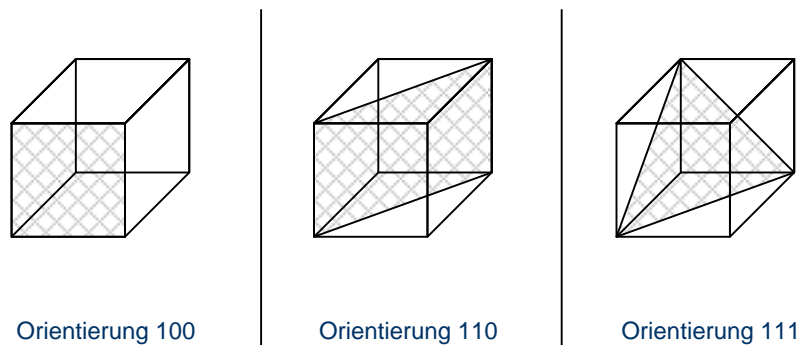


Abb. 2.2: Kristallorientierungen

Auch in der Mikromechanik ist die Kristallorientierung von besonderer Bedeutung. So lassen sich auf (110)-Silicium Mikrokanäle mit senkrechten Wänden herstellen, wohingegen sich bei (100)-Orientierung Flankenwinkel von $54,74^\circ$ ergeben.

2.3.2 Kristallziehverfahren nach Czochralski

Das polykristalline Silicium, wie es nach der Zonenreinigung vorliegt, wird in einem Quarztiegel knapp über dem Schmelzpunkt von Silicium verflüssigt. Der Schmelze können jetzt Dotierstoffe (z.B. Bor oder Phosphor) hinzugefügt werden, um entsprechende elektrische Eigenschaften des Einkristalls zu erzielen.

Ein Impfkristall (Einkristall) an einem drehbaren Stab wird an die Oberfläche der Siliciumschmelze gebracht. Dieser Keim gibt die Orientierung des Kristalls vor. Beim Kontakt des Keims mit der Schmelze lagert sich Silicium am Keim an und übernimmt dessen Kristallstruktur. Dadurch, dass die Tiegeltemperatur nur wenig über dem Schmelzpunkt von Silicium liegt, erstarrt das angelagerte Silicium sofort am Keimling und der Kristall wächst.

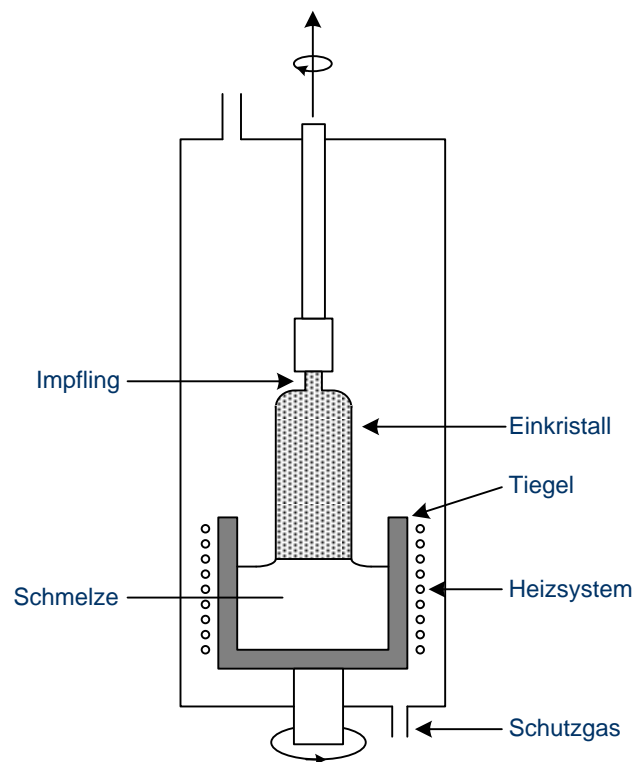


Abb. 2.3: Kristallziehen nach Czochralski

Der Keim wird unter ständigem Drehen langsam nach oben gezogen, wobei stetiger Kontakt zur Schmelze besteht. Der Tiegel dreht sich dabei entgegengesetzt zum Impfkristall. Eine konstante Temperierung der Schmelze ist unerlässlich, um ein gleich-

mäßiges Wachstum zu gewährleisten. Der Durchmesser des Einkristalls wird durch die Ziehgeschwindigkeit, die 2-25 cm/h beträgt, bestimmt. Je schneller gezogen wird, desto dünner wird der Kristall. Die gesamte Apparatur befindet dabei sich in einer Schutzgasatmosphäre, damit es nicht zu einer Oxidation des Siliciums kommt.

Der Nachteil dieses Verfahrens ist jedoch, dass sich die Schmelze während des Vorgangs immer mehr mit Dotierstoffen anreichert (bessere Löslichkeit der Dotierstoffe in der Schmelze als im Festkörper), so dass die Dotierstoffkonzentration entlang eines Siliciumstabes nicht konstant ist. Zudem können sich Verunreinigungen oder Metalle aus dem Tiegel lösen.

Die Vorteile dieses Verfahrens sind die geringen Kosten, und die Möglichkeit größere Wafer als beim Zonenziehen herzustellen.

2.3.3 Tiegelfreies Zonenziehen

Im Gegensatz zum Kristallziehverfahren nach Czochralski wird beim tiegelfreien Zonenziehen nicht das gesamte Polysilicium geschmolzen, sondern, wie bei der Zonenreinigung, nur ein kleiner Bereich (wenige Millimeter).

Auch hier sorgt ein Impfkristall, der an das Ende des polykristallinen Siliciumstabs herangeführt wird, zur Vorgabe der Kristallstruktur. Der Kristall wird aufgeschmolzen und übernimmt die Struktur des Keimlings. Die Heizzone wird langsam am Stab entlang geführt, der polykristalline Siliciumstab wandelt sich langsam in einen Einkristall um.

Da nur ein kleiner Bereich des polykristallinen Siliciums aufgeschmolzen wird, können sich kaum Verunreinigungen anlagern (wie bei der Zonenreinigung verlagern sich die Fremdatome ans Ende des Siliciumstabs). Die Dotierung erfolgt hier durch Zusätze der Dotierstoffe ins Schutzgas (z.B. mit Diboran oder Phosphin), welches die Apparatur umströmt.

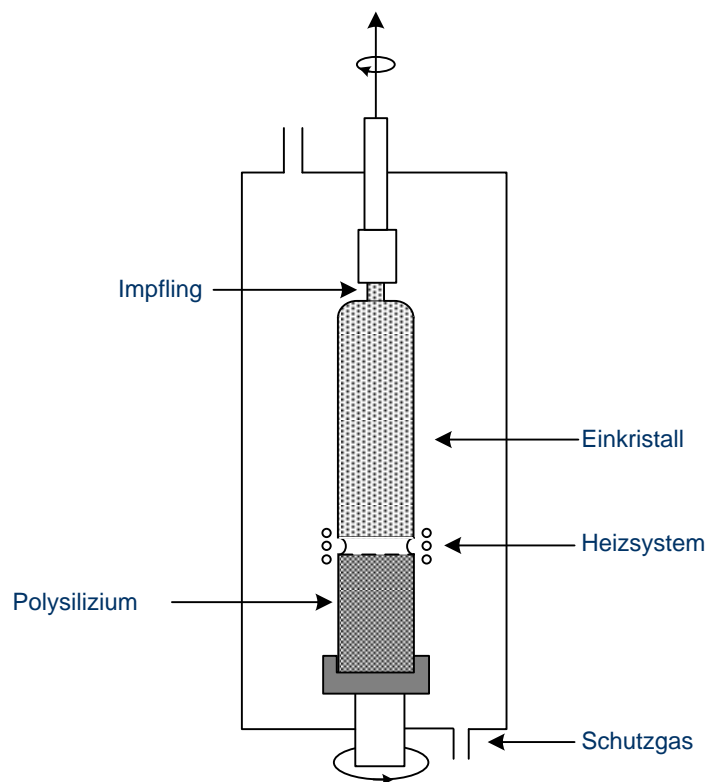


Abb. 2.4: Tiegelfreies Zonenziehen

2.4 Entstehung der Wafer

2.4.1 Wafervereinzelung und Oberflächenveredelung

Der Einkristallstab wird zunächst auf den gewünschten Durchmesser abgedreht und bekommt dann, je nach Kristallorientierung und Dotierung, einen oder zwei Flats. Der größere Flat dient dazu, die Wafer in der Fertigung exakt ausrichten zu können. Der zweite Flat dient zur Erkennung des Scheibentyps (Kristallorientierung, p-/n-Dotierung), ist aber nicht immer vorhanden. Bei Wafern ab 200 mm Durchmesser werden an Stelle der Flats sogenannte Notches verwendet. Dabei handelt es sich um eine winzige Einkerbungen am Scheibenrand, die ebenfalls eine Ausrichtung der Wafer ermöglicht, aber sehr viel weniger kostbare Fläche des Wafers beansprucht.

Sägen:

Mit einer Innenlochsäge, deren Schnittkante mit Diamantsplittern besetzt ist, wird der Einkristallstab in dünne Scheiben zersägt. Die Innenlochsäge bietet eine hohe Genau-

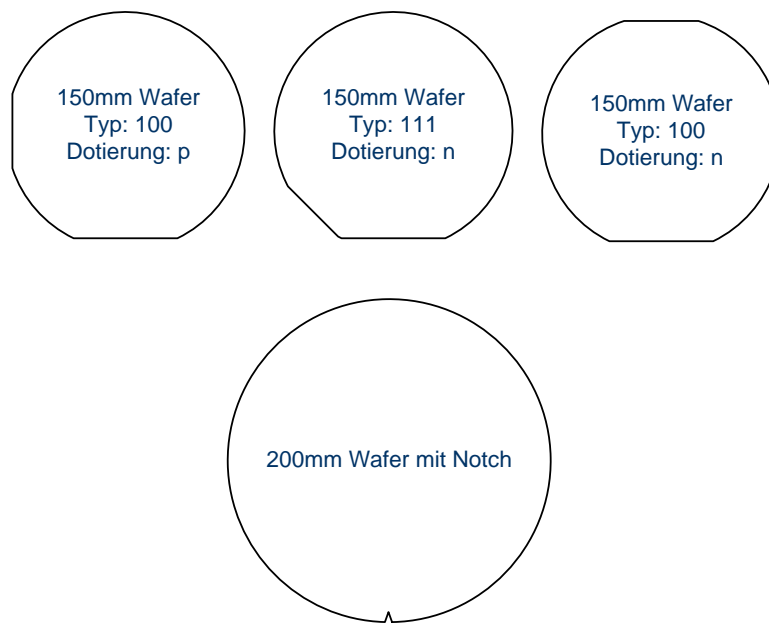


Abb. 2.5: Verschiedene Wafertypen

igkeit beim Sägen ohne Unebenheiten hervorzurufen. Bis zu 20 % des Kristallstabs gehen beim Sägen der Scheiben auf Grund der Dicke des Sägeblatts verloren. Nach dem Sägen besitzen die Scheiben eine aufgeraute Oberfläche und, auf Grund der mechanischen Belastung, Gitterschäden im Kristall. Zum Veredeln der Oberfläche durchlaufen die Scheiben mehrere Prozessschritte.

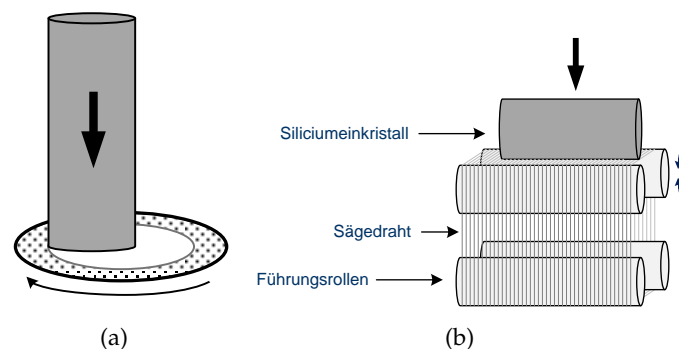


Abb. 2.6: Innenloch- und Drahtsäge

Läppen:

Mit körnigen Schleifmitteln (z.B. Aluminiumoxid) werden 50 μm (0,05 mm) der Scheibenoberfläche auf einer rotierenden Stahlscheibe abgetragen. Die Körnung wird dabei stufenweise verringert, jedoch wird die Oberfläche auf Grund der mechanischen Be-

handlung erneut geschädigt. Die Ebenheit nach dem Läppen beträgt ca. $2\text{ }\mu\text{m}$.

Scheibenrand abrunden:

In späteren Prozessen dürfen die Scheiben keine scharfen Kanten besitzen, da aufgetragene Schichten ansonsten abplatzen können. Dazu wird der Rand der Scheiben mit einem Diamantfräser abgerundet.

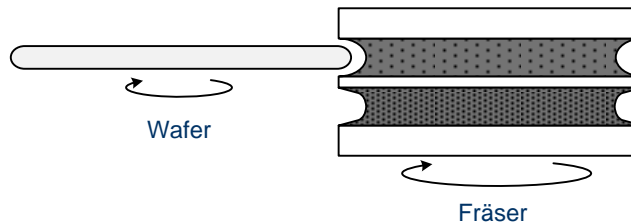


Abb. 2.7: Kantenverrundung

Ätzen:

In einem Tauchätzschritt, mit einer Mischung aus Fluss-, Essig- und Salpetersäure, werden noch einmal $50\text{ }\mu\text{m}$ abgetragen. Da es sich hierbei um einen chemischen Vorgang handelt, wird die Oberfläche nicht geschädigt. Kristallfehler werden endgültig behoben.

Polieren:

Dies ist der letzte Schritt zum fertigen Wafer. Am Ende des Polierschrittes besitzen die Wafer noch eine Unebenheit von weniger als 3 nm ($0,000003\text{ mm}$). Dazu werden die Scheiben mit einem Gemisch aus Natronlauge NaOH , Wasser und Siliciumdioxidkörnern behandelt. Das Siliciumdioxid entfernt weitere $5\text{ }\mu\text{m}$ von der Scheibenoberfläche, Die Natronlauge entfernt Oxid und beseitigt Bearbeitungsspuren der Siliciumdioxidkörner.

2.4.2 Geschichtliche Entwicklung der Wafergröße

Die Herstellung von integrierten Schaltkreisen auf Siliciumwafern begann Mitte der 1960-er Jahre auf Scheiben, mit einem Durchmesser von 25 mm . Heutzutage werden in modernen Halbleiterfertigungen Wafer mit einem Durchmesser von $150\text{--}300\text{ mm}$ eingesetzt. Bis 2012 rechnet man mit der Serienfertigung von Mikrochips auf Wafern mit einem Durchmesser von 450 mm , von denen zumindest schon Prototypen hergestellt wurden. Die Waferfläche beträgt dann mehr als das 300-fache der winzigen

1-Zoll-Wafer vor gut 50 Jahren; im Vergleich dazu nahm der Scheibendurchmesser nur um das 18-fache zu.

Typ	Durchmesser [mm]	Dicke [μm]	1. Flat [mm]	Durchbiegung [μm]
6 Zoll	$150 \pm 0,5$	≈ 700	55-60	25
8 Zoll	$200 \pm 0,5$	≈ 700	55-60	25
12 Zoll	$300 \pm 0,5$	≈ 700	55-60	25

Tab. 2.1: Typische Daten von Wafern

Mit größeren Wafern steigt die Durchsatzrate in der Herstellung von Chips erheblich, wodurch die Kosten in der Fertigung entsprechend gesenkt werden. So können bei identischer Strukturgröße mehr als doppelt so viele Chips auf einem 300-mm-Wafer hergestellt werden wie auf einem 200-mm-Wafer.

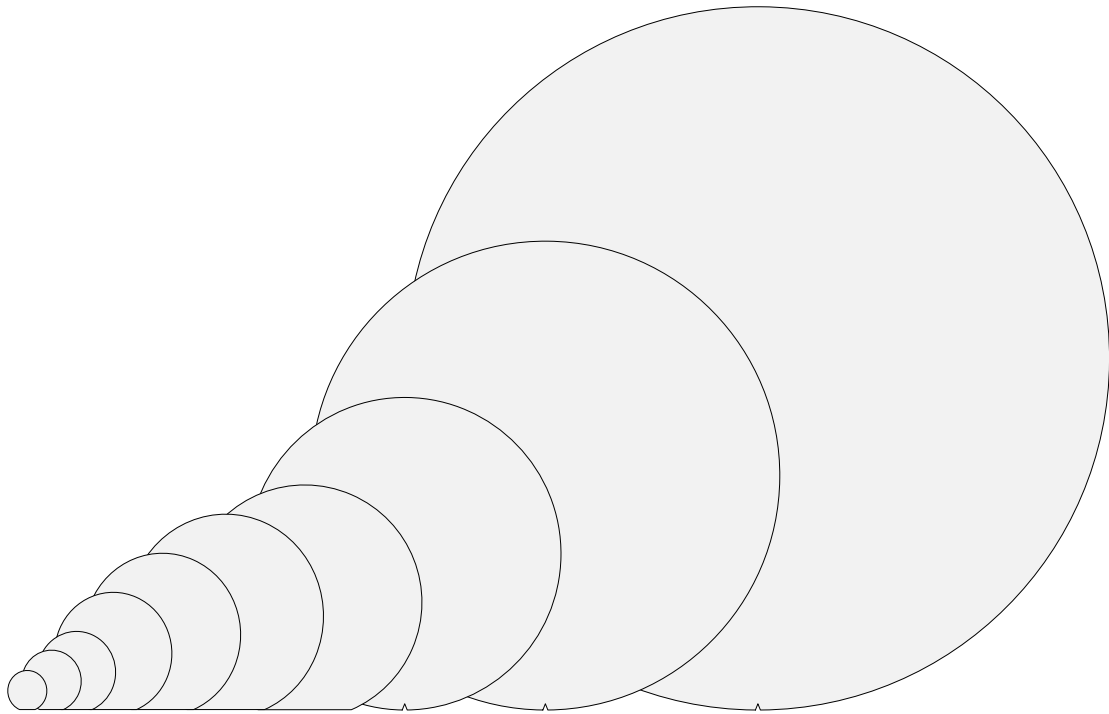


Abb. 2.8: Wafergrößen in der Übersicht: 25, 38, 51, 75, 100, 125, 150, 200, 300, 450 [mm] (Maßstabsgetreu)

2.5 Dotiertechniken

2.5.1 Begriff

Dotieren bedeutet das Einbringen eines Fremdstoffs in den Halbleiterkristall zur gezielten Änderung der Leitfähigkeit durch Elektronenmangel oder -überschuss. Im Gegensatz zum Dotieren bei der Waferherstellung selbst, wo der ganze Wafer dotiert wird, ermöglichen die beschriebenen Dotiertechniken das partielle Dotieren der Siliciumscheiben. Das Einbringen des Fremdstoffs kann mit verschiedenen Methoden erreicht werden: Diffusion, Implantation (und Legierung).

2.5.2 Diffusion

Diffusion bedeutet das selbständige Ausbreiten eines Stoffes in einem anderen Stoff auf Grund eines Konzentrationsunterschiedes, so verteilt sich z.B. ein Tropfen Tinte in einem Wasserglas nach einer bestimmten Zeit gleichmäßig. Im Siliciumkristall finden wir ein festes Gitter von Atomen vor, durch das sich der Dotierstoff bewegen muss. Das kann auf drei Arten geschehen:

- **Leerstellendiffusion:** Die Fremdatome besetzen leere Stellen im Kristallgitter die immer auftreten können. Ähnlich einer Löcherleitung entsteht ein Wechselspiel von besetzten Gitterplätzen und Lücken.
- **Zwischengitterdiffusion:** Die Fremdatome bewegen sich zwischen den Siliciumatomen im Kristallgitter hindurch.
- **Platzwechsel:** Fremdatome die sich im Kristallgitter befinden tauschen mit Siliciumatomen den Gitterplatz.

Der Dotierstoff kann sich so lange im Halbleiterkristall bewegen, bis entweder ein Konzentrationsgefälle ausgeglichen ist, oder die Temperatur so weit gesenkt wurde, dass die Atome sich nicht mehr bewegen können. Die Geschwindigkeit des Diffusionsvorgangs hängt von mehreren Faktoren ab:

- Dotierstoff
- Konzentrationsunterschied
- Temperatur

- Substrat
- Kristallorientierung des Substrats

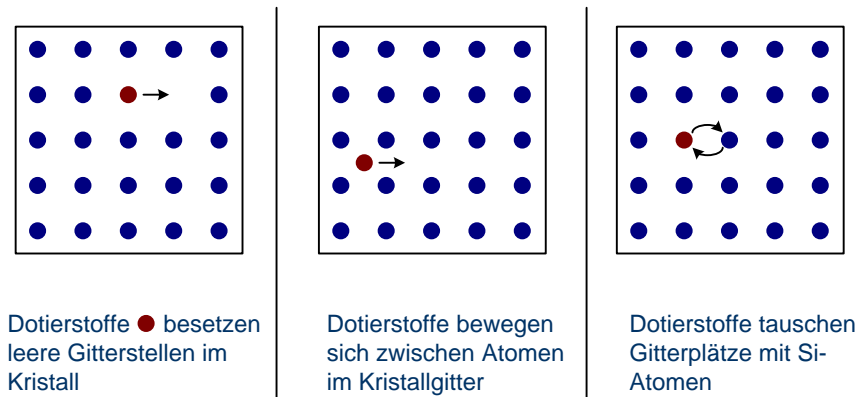


Abb. 2.9: Diffusionsvorgänge im Kristallgitter

Diffusion mit erschöpflicher Quelle:

Diffusion mit einer erschöpflichen Quelle bedeutet, dass der Dotierstoff nur begrenzt zur Verfügung steht. Je länger der Diffusionsprozess andauert, umso geringer wird die Konzentration an der Oberfläche; dafür steigt die Eindringtiefe in das Substrat. Der Diffusionskoeffizient eines Stoffes gibt dabei an, wie schnell er sich im Kristall bewegt. Arsen mit einem geringen Diffusionskoeffizienten dringt langsamer in das Substrat ein, als bspw. Phosphor oder Bor.

Diffusion mit unerschöpflicher Quelle:

Der Dotierstoff steht hier unbegrenzt zur Verfügung \Rightarrow die Konzentration an der Oberfläche bleibt während des Vorgangs konstant, da Teilchen, die in das Substrat eingedrungen sind, fortwährend nachgeliefert werden.

2.5.3 Diffusionsverfahren

Bei den nachfolgenden Prozessen befinden sich die Wafer in einem Quarzrohr, das über die gesamte Länge auf eine bestimmte Temperatur aufgeheizt wird.

Diffusion aus der Gasphase:

Ein Trägergas (Stickstoff, Argon) wird mit dem gewünschten Dotierstoff (ebenfalls in Gasform, z.B. Phosphin: PH_3 oder Diboran: B_2H_6) angereichert und über die Siliciumscheiben geleitet, wo der Konzentrationsausgleich stattfinden kann.

Feststoffdiffusion:

Zwischen den Wafern befinden sich Scheiben, auf denen der Dotierstoff aufgebracht wurde. Steigt die Temperatur im Quarzrohr, diffundiert von den Quellscheiben der Dotierstoff in die Atmosphäre aus. Über ein Trägergas wird der Dotierstoff anschließend gleichmäßig im Quarzrohr verteilt und gelangt so an die Oberfläche der Wafer.

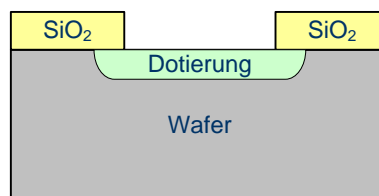
Diffusion mit flüssiger Quelle:

Als flüssige Quelle dienen Borbromid BBr_3 oder Phosphoroxychlorid POCl_3 . Ein Trägergas wird durch die Flüssigkeit geleitet und transportiert den Dotierstoff so zu den Wafern, wo er dann an die Oberfläche der Siliciumscheiben gelangt.

Da nicht der gesamte Wafer dotiert werden soll, werden bestimmte Bereiche mit Siliciumdioxid maskiert. Das Oxid können die Dotierstoffe nicht durchdringen, weshalb an diesen Stellen keine Dotierung stattfindet. Um Spannungen oder gar Brüche der Scheiben zu vermeiden wird das Rohr schrittweise ($10\text{ }^\circ\text{C pro Minute}$) auf ca. $900\text{ }^\circ\text{C}$ aufgeheizt, wo der Dotierstoff dann zu den Scheiben geleitet wird. Um den Diffusionsvorgang in Gang zu setzen, wird die Temperatur anschließend auf ca. $1200\text{ }^\circ\text{C}$ erhöht.

Charakteristik:

- Da viele Wafer gleichzeitig bearbeitet werden können ist dieses Verfahren recht günstig
- Befinden sich bereits Fremdstoffe früherer Dotierungen im Kristall können diese bei einer erneuten Temperaturbelastung ausdiffundieren
- Im Quarzrohr lagern sich mit der Zeit Dotierstoffe an, die bei nachfolgenden Dotierungen zusätzlich vom Trägergas zu den Wafern transportiert werden
- Dotierstoffe breiten sich im Kristall nicht nur senkrecht, sondern auch seitlich aus, so dass die Dotierfenster immer großflächiger dotiert werden als gewünscht



Der Dotierstoff wandert auch
seitlich unter die Maskierung

Abb. 2.10: Diffusion mit Oxidmaske

2.5.4 Ionenimplantation

Bei der Ionenimplantation werden geladene Dotierstoffe (Ionen) in einem elektrischen Feld beschleunigt und auf die Wafer gelenkt. Die Eindringtiefe lässt sich sehr genau festlegen, indem die zur Beschleunigung der Ionen benötigte Spannung verringert oder erhöht wird. Da der Prozess bei Raumtemperatur stattfindet, können vorher eingebrachte Dotierungen nicht ausdiffundieren. Wie bei der Diffusion werden Stellen, die nicht dotiert werden sollen, mit einer Maske verdeckt, wobei bei der Implantation eine Maskierung aus Fotolack ausreicht. Ein Implanter besteht aus folgenden Komponenten:

- **Ionenquelle:** Das Dotiergas (z.B. Bortrifluorid BF_3) wird ionisiert (Elektronen werden von einer Glühkathode emittiert und stoßen mit den Gasteilchen zusammen. Durch Stoßionisation werden stets positive Ionen und freie Elektronen erzeugt)
- **Vorbeschleuniger:** Die Ionen werden mit ca. 30 Kiloelektronenvolt aus der Ionenquelle gezogen
- **Massenseparator:** Die geladenen Teilchen werden durch ein Magnetfeld um 90° abgelenkt. Zu leichte/schwere Teilchen werden mehr/weniger abgelenkt als die gewünschten Ionen und mit Blenden hinter dem Separator abgefangen
- **Beschleunigungsstrecke:** Mit mehreren 100 keV werden die Teilchen auf ihre Endenergie beschleunigt (200 keV beschleunigen Borionen auf ca. 2.000.000 m/s)
- **Linsen:** Über das gesamte System sind Linsen verteilt, die den Ionenstrahl fokussieren
- **Ablenkungsvorrichtungen:** Kondensatoren lenken die Ionen auf die gewünschte Stelle
- **Waferstation:** Die Wafer werden entweder einzeln oder auf großen rotierenden Rädern in den Ionenstrahl gebracht und bestrahlt

Eindringtiefe von Ionen im Wafer:

Im Gegensatz zur Diffusion dringen die Teilchen nicht auf Grund ihrer Eigenbewegung ein, sondern werden mit hoher Geschwindigkeit in das Kristallgitter geschossen. Dabei werden sie durch Zusammenstöße mit den Siliciumatomen abgebremst. Durch den Aufprall werden die Siliciumatome von ihren Gitterplätzen gestoßen, die Dotie-

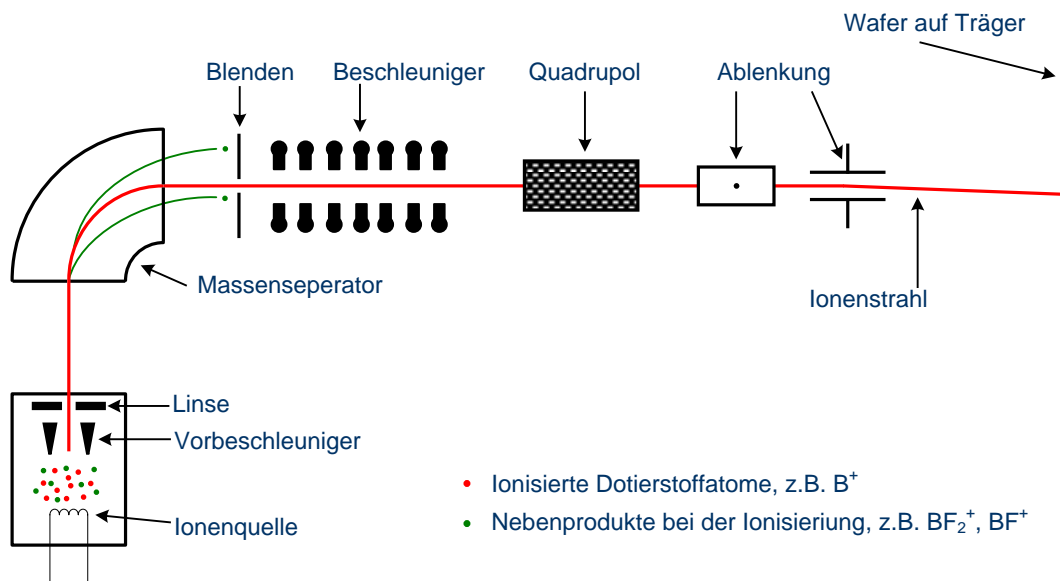


Abb. 2.11: Darstellung einer Implantationsanlage

rionen selbst lagern sich meist auf Zwischengitterplätzen an. Dort sind sie elektrisch nicht aktiv, da keine Bindungen mit anderen Atomen vorliegen, die freie Ladungsträger hervorrufen könnten. Die verschobenen Siliciumatome müssen wieder ins Kristallgitter eingebaut, und die elektrisch nicht aktiven Dotierstoffe aktiviert werden.

Ausheilen des Kristallgitters und Aktivierung der Dotierstoffe:

Durch einen Temperaturschritt bei ca. 1000 °C werden die Dotierstoffe auf Gitterplätze bewegt (vorher befinden sich nur ca. 5 % der Dotieratome auf Gitterplätzen). Die Gitterschäden durch die Zusammenstöße werden bereits bei ca. 500 °C ausgeheilt. Da sich die Dotieratome während den hohen Temperaturen im Substrat bewegen, werden diese Schritte nur sehr kurze Zeit durchgeführt.

Channeling:

Das verwendete Substrat liegt als Einkristall vor, d.h. die Siliciumatome sind regelmäßig angeordnet und bilden Kanäle. Die eingeschossenen Dotierstoffatome verlaufen dann parallel zu diesen Kanälen, werden nur schwach abgebremst und dringen sehr tief in das Substrat ein. Um dies zu verhindern gibt es unterschiedliche Möglichkeiten:

- **Waferausrichtung:** Die Wafer werden um ca. 7° zur Strahlrichtung ausgelenkt. Dadurch werden die Ionen nicht parallel zu den Gitterkanälen eingeschossen und durch Zusammenstöße frühzeitig abgebremst.

- **Streuoxid:** Auf der Waferoberfläche wird ein dünnes Oxid aufgebracht, das die Ionen ablenkt und ein senkrechtres Eintreffen im Substrat verhindert

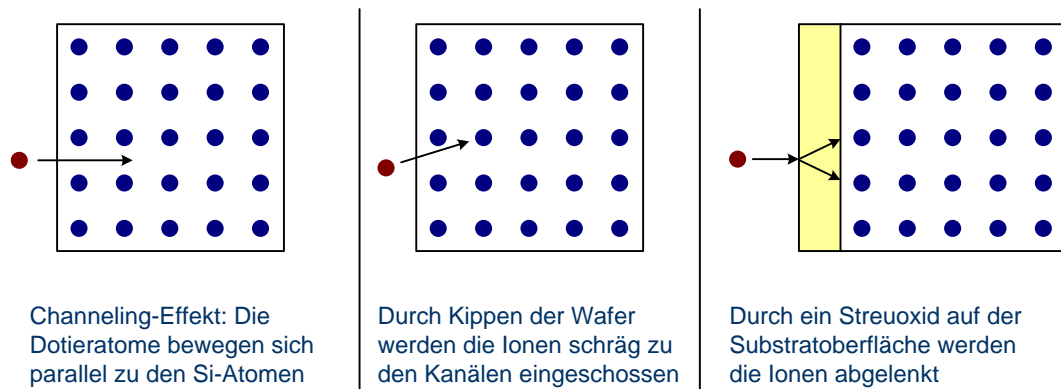


Abb. 2.12: Channelingeffekt im Einkristallgitter

Charakteristik:

- Die Reproduzierbarkeit der Ionenimplantation ist sehr hoch
- Der Ablauf bei Raumtemperatur verhindert ein Ausdiffundieren anderer Dotierstoffe
- Als Maske dient Fotolack, Oxid wie bei der Diffusion ist nicht nötig
- Ionenimplanter sind teuer, die Kosten pro bearbeiteter Scheibe sind hoch
- Die Dotierstoffe breiten sich nicht seitlich unter der Maskierung aus (nur minimal durch Zusammenstöße)
- Nahezu jedes Element kann in höchster Reinheit implantiert werden
- Ähnlich den Ablagerungen von Dotierstoffen bei der Diffusion im Quarzrohr, können sich an Wänden oder Blenden Ionen ablagern, die bei späteren Implantationen abgelöst werden und auf die Wafer gelangen
- Dreidimensionale Strukturen (z.B. Gräben) können mit der Ionenimplantation nicht dotiert werden
- Der Implantationsprozess findet unter Hochvakuum statt, welches man mit mehreren Turbomolekular- oder Kryopumpen erzeugen kann

Es gibt verschiedene Implantertypen, wobei meistens Mittel- und Hochstromimplanter zum Einsatz kommen. Mittelstromimplanter sind für kleine bis mittlere Dosen an

Ionen geeignet ($10^{11} - 10^{15}$ Ionen/cm²), Hochstromimplanter für Dosen von $10^{15} - 10^{17}$ Ionen/cm². Die Ionenimplantation hat sich auf Grund der Vorteile gegenüber der Diffusion größtenteils durchgesetzt.

Dotieren mittels Legierung

Der Vollständigkeit halber sei noch erwähnt, dass es neben diesen beiden Verfahren noch die Dotierung mittels Legierung gibt. Da dieses Verfahren aber Nachteile, wie z.B. Rissbildung im Substrat mit sich bringt, wird es in der heutigen Halbleitertechnik kaum mehr eingesetzt.

3 Oxidation

3.1 Übersicht

3.1.1 Anwendungen

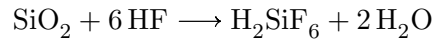
Oxid findet in der Halbleiterfertigung Anwendung in den verschiedensten Gebieten:

- zur Isolation (z.B. zwischen Metallisierungsschichten)
- als Streuschicht (z.B. Ionenimplantation)
- als Anpassungsschicht
- zur Planarisierung (z.B. um Kanten zu entschärfen)
- als Maskierschicht (z.B. Diffusion)
- als Justiermarken (Justierpunkte in der Fototechnik)
- als Schutzschicht (vor mechanischer Beschädigung)

3.1.2 Eigenschaften von Oxid

In Verbindung mit Silicium tritt Oxid als Siliciumdioxid SiO_2 auf. Es lässt sich auf dem Wafer in sehr dünnen Schichten sehr gleichmäßig herstellen.

SiO_2 ist sehr widerstandsfähig und kann in der Halbleiterfertigung nur durch Flusssäure HF nasschemisch geätzt werden. Wasser und andere Säuren greifen das Oxid nicht an, wegen der Kontaminationsgefahr durch Metallionen können Alkalilaugen (KOH, NaOH u.a.) nicht eingesetzt werden. Diese spielen jedoch in der Mikromechanik eine Rolle beim anisotropen Ätzen.



Des Weiteren bietet sich Oxid zur Funktion von Schaltungen an, da es die unterschiedlichsten elektrischen Anforderungen erfüllt (z.B. als Gate- oder Feldoxid).

3.2 Erzeugung von Oxidschichten

3.2.1 Thermische Oxidation

Allgemein:

Bei der thermischen Oxidation werden die Siliciumwafer bei ca. 1000 °C in einem Oxidationsofen oxidiert. Dieser Ofen besteht im Wesentlichen aus einem Quarzrohr in dem sich die Wafer auf einem Carrier aus Quarzglas befinden, drei bis fünf getrennt regelbaren Heizwicklungen und verschiedenen Gaszuleitungen. Das Quarzglas hat einen sehr hohen Schmelzpunkt (weit über 1500 °C) und ist deshalb sehr gut für Hochtemperaturprozesse geeignet. Damit es nicht zu Scheibenverzug oder Scheibensprüngen kommt, wird das Quarzrohr in sehr kleinen Schritten (max. 10 °C pro Minute) aufgeheizt. Die Temperierung ist mittels der getrennten Heizwicklungen im gesamten Rohr über eine Länge von ca. 1 m auf $\pm 0,5$ °C exakt regelbar.

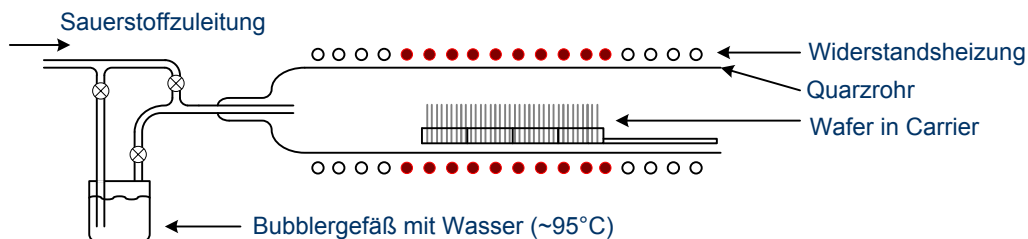
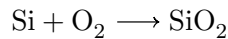


Abb. 3.1: Darstellung eines Oxidationsofens

Der Sauerstoff strömt dann als Gas über die Wafer und reagiert an der Oberfläche zu Siliciumdioxid. Es entsteht eine glasartige Schicht mit amorpher Struktur. Je nach Prozessgas finden dann verschiedene Oxidationen statt (eine thermische Oxidation muss naturgemäß auf einer Siliciumoberfläche stattfinden). Die thermische Oxidation unterteilt sich in die trockene und feuchte Oxidation, welche sich wiederum in die nasse Oxidation und die H_2O_2 -Verbrennung gliedern lässt.

Trockene Oxidation:

Der Oxidationsprozess findet unter reiner Sauerstoffatmosphäre statt. Dabei reagiert Silicium mit Oxid zu Siliciumdioxid:



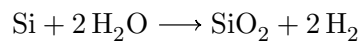
Dieser Prozess findet in der Regel bei 1000-1200 °C statt. Zur Erzeugung von sehr stabilen und dünnen Oxiden wird die Oxidation bei ca. 800 °C durchgeführt.

Eigenschaften der Trockenoxidation:

- langsames Oxidwachstum
- hohe Dichte
- hohe Durchbruchspannung (für elektrisch stark beanspruchte Oxide, z.B. Gateoxid)

Nasse Oxidation:

Bei der nassen Oxidation wird der Sauerstoff durch ein Bubbler-Gefäß mit Wasser (ca. 95 °C) geleitet, so dass sich zusätzlich zum Sauerstoff auch Wasser in Form von Wasserdampf im Quarzrohr befindet. Daraus ergibt sich folgende Reaktionsgleichung:



Dieser Prozess findet bei 900-1000 °C statt. Die nasse Oxidation weist folgende Eigenschaften auf:

- hohes Oxidwachstum schon bei geringer Temperatur (Verwendung von Maskierschichten, Feldoxid)
- geringere Qualität als Trockenoxid

H₂O₂-Verbrennung:

Bei der H₂O₂-Verbrennung wird neben hochreinem Sauerstoff auch hochreiner Wasserstoff verwendet. Die beiden Gase werden getrennt in das Quarzrohr geleitet und an der Eintrittsöffnung verbrannt. Damit es nicht zu einer Knallgasreaktion mit dem

Temperatur	Trockene Oxidation	Nasse Oxidation
900 °C	19 nm/h	100 nm/h
1000 °C	50 nm/h	400 nm/h
1100 °C	120 nm/h	630 nm/h

Tab. 3.1: Vergleich der Aufwachsrate bei trockener und nasser Oxidation

hochbrennbaren Wasserstoff kommt, muss die Temperatur über 500 °C liegen, die Gase reagieren dann in einer stillen Verbrennung. Dieses Verfahren ermöglicht die Erzeugung von schnell wachsenden und nur wenig verunreinigten Oxidschichten. Damit lassen sich sowohl dicke Oxide, als auch dünne Schichten bei vergleichsweise geringer Temperatur (900 °C) herstellen. Die niedrige Temperatur erlaubt auch die thermische Belastung von bereits dotierten Wafern.

Bei allen thermischen Oxidationen ist das Oxidwachstum auf 111-orientierten Substraten höher als auf 100-orientierten. Außerdem erhöht ein sehr hoher Anteil an Dotierstoffen im Substrat das Wachstum deutlich.

Ablauf des Oxidationsvorgangs:

Zu Beginn reagiert der Sauerstoff an der Waferoberfläche zu Siliciumdioxid. Nun befindet sich eine Oxidschicht auf dem Substrat durch die der Sauerstoff zunächst diffundieren muss, um mit dem Silicium reagieren zu können. Die Aufwachsrate hängt nur zu Beginn von der Reaktionszeit zwischen Silicium und Oxid ab; ab einer gewissen Dicke wird die Oxidationsgeschwindigkeit von der Diffusionsgeschwindigkeit des Oxids durch das Siliciumdioxid bestimmt. Mit zunehmender Oxiddicke verlangsamt sich also das Wachstum. Da die entstandene Schicht amorph ist, sind nicht alle Bindungen der Siliciumatome intakt; es gibt teilweise freie Bindungen (freie Elektronen und Löcher) an der Si–SiO₂-Grenzschicht. So ergibt sich an diesem Übergang insgesamt eine leicht positive Ladung. Da sich diese Ladung negativ auf Bauteile auswirken kann wird versucht sie so gering wie möglich zu halten. Das kann beispielsweise mit höherer Oxidationstemperatur erreicht werden, oder durch Verwendung der nassen Oxidation, die ebenfalls nur eine sehr geringe Ladung verursacht.

Segregation:

Bei der thermischen Oxidation wird Silicium, durch die Reaktion mit Sauerstoff zu Siliciumdioxid, verbraucht. Das Verhältnis der aufgewachsenen Oxidschicht zu verbrauchtem Silicium beträgt 2,27; d.h. das Oxid wächst zu 45 % der Oxiddicke in das

Substrat ein.

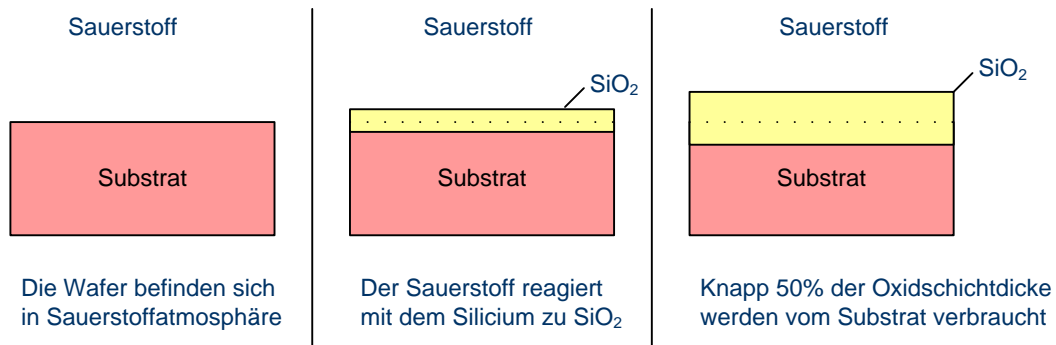


Abb. 3.2: Aufwuchsverhalten von Oxid auf Silicium

Dotierstoffe die sich im Substrat befinden, können im Siliciumkristall oder im Oxid eingebaut werden, dies hängt davon ab, in welchem Material sich der Dotierstoff besser löst. Das Verhalten lässt sich mit einer Formel berechnen, k wird dabei als Segregationskoeffizient bezeichnet:

$$k = \frac{\text{Löslichkeit des Dotierstoffs in Silicium}}{\text{Löslichkeit des Dotierstoffs in SiO}_2}$$

Ist k größer 1 werden die Dotierstoffe an der Oberfläche des Substrats eingebaut, bei k kleiner 1 reichern sich die Dotierstoffe im Oxid an.

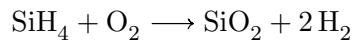
3.2.2 Oxidation durch Abscheidung

Bei der thermischen Oxidation wird Silicium des Wafers zur Oxidbildung verbraucht. Ist die Silicioberfläche jedoch durch andere Schichten verdeckt, muss man das Oxid über Abscheidungsverfahren aufbringen, bei denen neben Sauerstoff auch Silicium selbst hinzugefügt wird. Die zwei wichtigsten Verfahren dabei sind die Silanpyrolyse und die TEOS-Abscheidung. Eine ausführliche Beschreibung der Verfahren folgt später im Kapitel Abscheidung.

Silanpyrolyse:

Pyrolyse bedeutet, dass chemische Verbindungen durch Wärme gespalten werden, in diesem Fall das Gas Silan SiH_4 und hochreiner Sauerstoff O_2 . Da sich das giftige Silan bei einer Konzentration von 3 % in der Umgebungsluft selbst entzündet muss es mit Stickstoff oder Argon auf 2 % verdünnt werden. Bei ca. 400 °C reagiert das Silan und

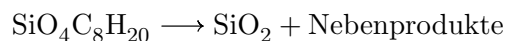
der Sauerstoff zu Siliciumdioxid und Wasserstoff:



Das Siliciumdioxid ist nur von geringer Qualität. Alternativ kann eine Hochfrequenzanregung über ein Plasma bei ca. 300 °C zur Dioxidabscheidung verwendet werden. So entsteht ein etwas stabileres Oxid.

TEOS-Abscheidung:

Das bei dieser Methode verwendete Tetraethylorthosilicat, kurz TEOS ($\text{SiO}_4\text{C}_8\text{H}_{20}$) enthält die beiden benötigten Elemente Silicium und Sauerstoff. Unter Vakuum geht die bei Raumtemperatur flüssige Verbindung bereits in Gasform über (durch das Vakuum herrscht ein geringerer Druck auf der Oberfläche der Flüssigkeit, die Siedetemperatur wird herabgesetzt). Das Gas wird in ein beheiztes Quarzrohr überführt und dort bei ca. 750 °C gespalten.



Das Siliciumdioxid scheidet sich auf den Wafern ab, die Nebenprodukte (z.B. H_2O - Wasserdampf) werden abgesaugt. Die Gleichmäßigkeit dieses Oxids wird durch den Druck im Quarzrohr und die Prozesstemperatur bestimmt. Es ist elektrisch stabil und sehr rein.

3.3 Die LOCOS-Technik

3.3.1 Höchstintegration auf Chips

In der Halbleitertechnik werden Strukturen mittels Belichtungs- und Ätzverfahren strukturiert. So entstehen Stufen, an denen sich z.B. Fotolack ansammeln kann und so das Auflösungsvermögen in der Fototechnik verringert. Durch die isotrope Ätzcharakteristik (die Ätzung geschieht sowohl vertikal als auch horizontal) müssen Lackmasken angepasst werden, damit unterätzte Strukturen die richtige Größe besitzen.

An diesen Stufen treten auch Probleme bei der Metallisierung auf, da die Leiterbahnen hier verengt werden, so dass Schäden durch Elektromigration die Folge sind.

Um eine hohe Packungsdichte zu erreichen, also möglichst viele Bauelemente auf möglichst geringer Fläche unterzubringen, müssen die Stufen und Unebenheiten vermieden werden. Dies ist beispielsweise mit der LOCOS-Technik realisierbar: LOCal Oxidation of Silicon (Lokale Oxidation von Silicium).

3.3.2 Der Vogelschnabel

Bei der LOCOS-Technik nutzt man die unterschiedlichen Oxidationsgeschwindigkeiten von Silicium und Siliciumnitrid zur lokalen Maskierung der Scheibenoberfläche aus.

Mit einer Siliciumnitridschicht maskiert man die Stellen an denen kein Oxid aufwachsen soll, es bildet sich nur eine Oxidschicht auf den Nitrid freien Bereichen. Da Silicium und Siliciumnitrid unterschiedliche Ausdehnungskoeffizienten besitzen wird eine dünne Schicht Oxid, das Padoxid, zwischen Nitridmaske und Substrat aufgebracht um Spannungen durch Temperaturänderungen zu vermeiden.

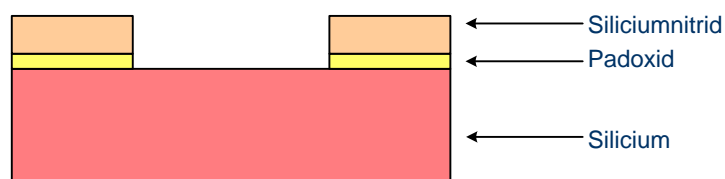


Abb. 3.3: Schichtaufbau vor dem LOCOS-Prozess

Zur seitlichen Isolation von Transistoren bringt man nun ein Feldoxid (FOX) auf der freien Siliciumoberfläche auf. Während sich bei der Feldoxidation auf dem Silicium eine Siliciumdioxidschicht bildet, verursacht das Padoxid eine seitliche Sauerstoffdiffusion unter die Nitridmaske und somit ein leichtes Oxidwachstum am Rand der Maskierung. Der Oxidaufläufer hat die Form eines Vogelschnabels, dessen Länge vom Oxidationsprozess, sowie von der Dicke des Nitrids und des Padoxids abhängt.

Neben diesem Effekt, der bis zu $1\ \mu\text{m}$ der Fläche für Bauelemente einnehmen kann, tritt bei einer feuchten Oxidation außerdem der so genannte White-Ribbon- oder Kooi-Effekt auf. Dabei reagiert Nitrid aus der Maskierschicht mit Wasserstoff zu Ammoniak (NH_3), der zur Siliciumoberfläche diffundiert und dort zu einer Nitridation führt.

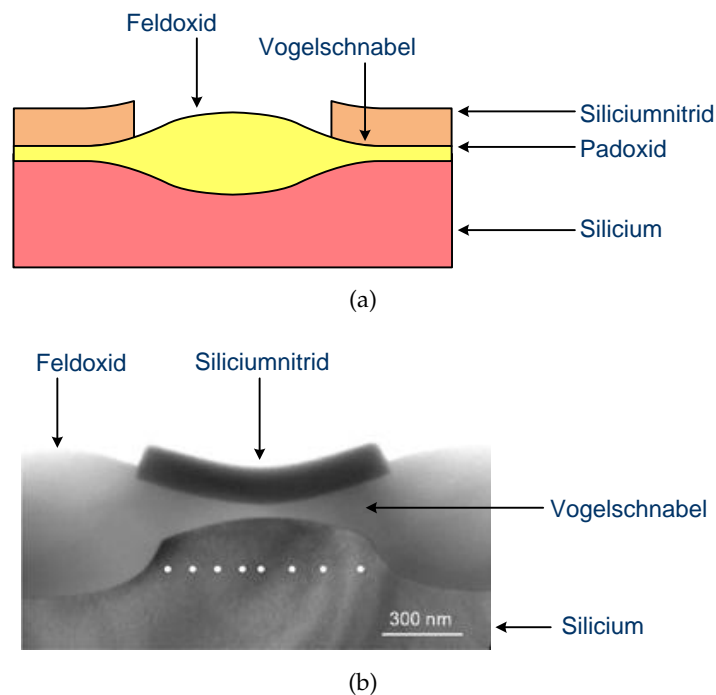


Abb. 3.4: LOCOS-Struktur nach der Oxidation

Vor der Gateoxidation muss dieses Nitrid entfernt werden, da es sonst als Maskierung wirkt.

Trotz dieser negativen Effekte ist die LOCOS-Technik ein geeignetes Verfahren um die hohe Packungsdichte zu ermöglichen. Durch Ätzen einer Oxidschicht wären so geringe Abmessungen für die Integration von Transistoren und anderen Bauteilen nicht erreichbar. Des Weiteren ist wegen der geringeren Unebenheit, ohne die Kanten- und Stufenbildung, die Auflösung in der Fototechnik verbessert. Das FOX lässt sich noch etwas zurückätzen, dadurch wird zwar das aufgewachsene Oxid leicht reduziert, die Länge des Vogelschnabels nimmt jedoch ab und die Oberfläche wird wiederum etwas mehr eingeebnet: fully recessed LOCOS.

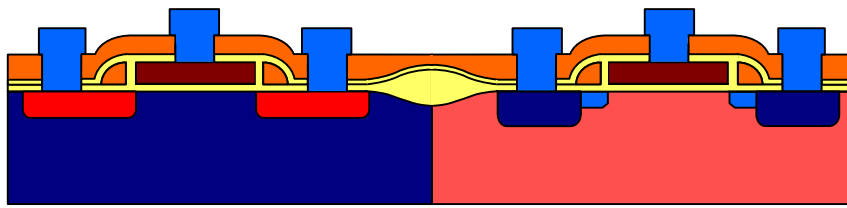


Abb. 3.5: Anwendungsbeispiel der LOCOS-Technik zur seitlichen Isolation zweier Transistoren

3.4 Schichtdickenmessung

3.4.1 Messtechnik

Oxidschichten sind lichtdurchlässige Schichten. Wird Licht auf den Wafer gestrahlt und reflektiert, ändern sich bestimmte Eigenschaften der Lichtwellen, die mit Messgeräten erfasst und ausgewertet werden können. Sollen mehrere übereinander liegende Schichten gemessen werden, müssen diese unterschiedliche optische Indizes haben, damit eine Unterscheidung der Materialien möglich ist.

Damit die Schichtdicke auf dem gesamten Wafer kontrolliert werden kann, werden mehrere Messpunkte (z.B. 5 Punkte bei 150 mm, 9 Punkte bei 200 mm, 13 Punkte bei 300 mm Wafern) gemessen. Dabei dürfen nicht nur die Werte der einzelnen Punkte mit dem vorgegebenen Sollwert verglichen werden, sondern auch die Dicke der Punkte untereinander, da auch die Homogenität wichtig für die weiteren Prozesse ist. Ist die aufgebrachte Schicht zu dick oder zu dünn, muss Material entfernt (z.B. durch Ätzen) oder zusätzliches (Wiederholung des entsprechenden Prozesses) aufgebracht werden.

3.4.2 Interferometrie

Bei der Überlagerung von Lichtwellen, können sich diese verstärken, abschwächen oder auch auslöschen. Dieses Phänomen nutzt man in der Halbleiterfertigung zur Messung von transparenten Schichten aus.

Auf den Wafer einfallende Lichtstrahlen werden an einer transparenten Schicht teilweise reflektiert, ein Teil der Strahlen durchdringt das Material aber auch und wird an der darunter befindlichen Schicht reflektiert.

Es wird dabei ein Spektrum an unterschiedlichen Wellenlängen auf den Wafer ge-

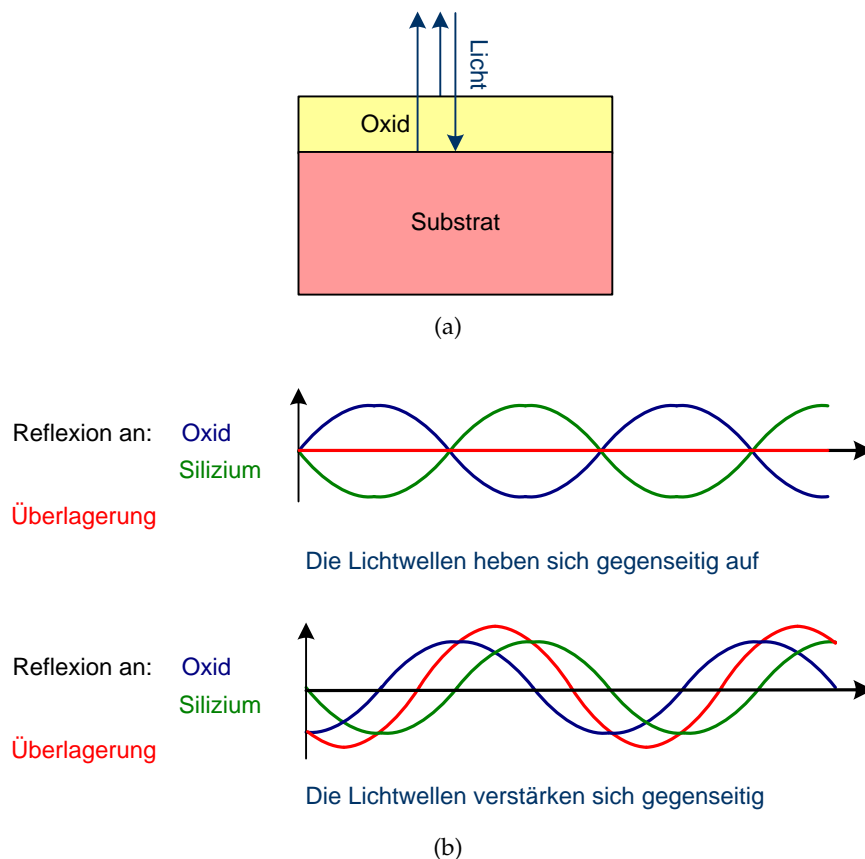


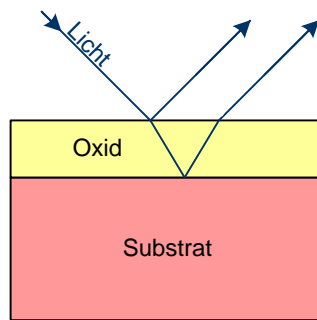
Abb. 3.6: (a) Prinzip der Interferenz, (b) destruktive und konstruktive Überlagerung von Lichtwellen

strahlt, und je nach Dicke der durchstrahlten Schicht überlagern sich die reflektierten Strahlen unterschiedlich und ergeben eine für jede Schichtdicke und jedes Material charakteristische Überlagerung. Mit Hilfe eines Photometers kann aus dem reflektierten Licht die Schichtdicke ermittelt werden.

Interferenzmessungen sind bei Schichten möglich, deren Dicke in etwa einem Viertel der eingestrahnten Lichtwellenlänge entspricht.

3.4.3 Ellipsometrie

Ellipsometrie ist die Bestimmung von optischen Eigenschaften durch Änderung der Polarisation von Licht. Linear polarisiertes Licht (die Lichtwellen haben eine bestimmte Schwingung) wird unter einem festen Winkel auf den Wafer gestrahlt.



Schräg eingestrahltes Licht wird gebrochen und reflektiert;
gleichzeitig ändert sich die Polarisierung

Abb. 3.7: Ellipsometrie

Bei der Reflexion an Waferoberfläche bzw. an der Grenzschicht zwischen zwei Schichten wird das Licht unpolarisiert. Diese Änderung kann dann mit einem Analysator gemessen werden. Aus den bekannten optischen Eigenschaften der Schichten (z.B. Brechungswinkeln, Absorptionskoeffizient u.a.), der eingestrahlten Wellenlänge und der Polarisierung kann die Schichtdicke bestimmt werden.

Im Gegensatz zur Messung mittels Interferenz, ist die Ellipsometrie für dünne Schichten geeignet, deren Dicke weniger als ein Viertel der eingestrahlten Lichtwellenlänge beträgt.

3.4.4 Beurteilung der Messung

Bei diesen optischen Messverfahren erfolgt die Schichtdickenbestimmung nur indirekt, die optischen Parameter der zu messenden Schichten müssen dazu hinreichend bekannt sein. Mit Hilfe dieser Werte wird dann ein Modell der Schichtabfolge auf dem Wafer erstellt und eine Messung simuliert. Das Ergebnis wird anschließend mit der tatsächlichen Messung verglichen und die Schichtdicken (aber auch andere optische Indizes) der im Modell hinterlegten Materialien variiert, bis Simulation und Messung bestmöglich übereinstimmen.

Je mehr Parameter im Modell variiert werden, desto leichter kann eine Übereinstimmung mit der Messung gefunden werden, aber desto unsicherer ist auch das Ergebnis. Meist gibt ein Parameter (goodness of fit, GOF; Anpassungsgüte) darüber Auskunft, wie gut die Simulation mit der Messung übereinstimmt (0-100 %).

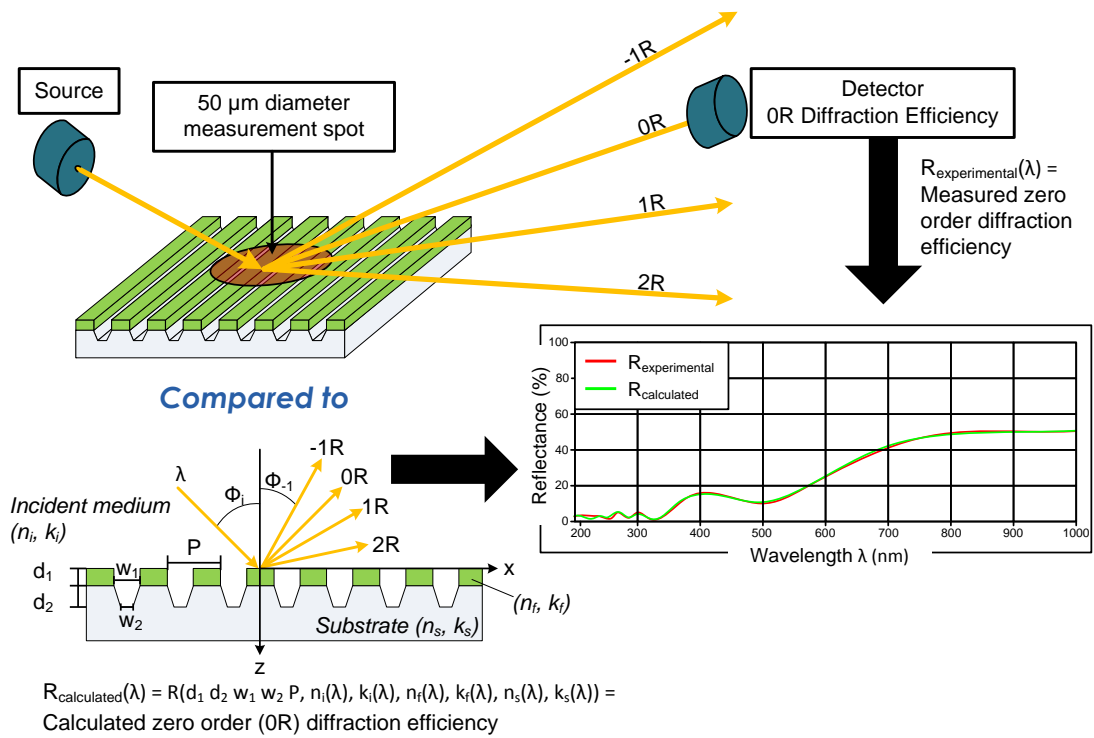


Abb. 3.8: Vergleich von Messung und Simulation

4 Abscheidung

4.1 Plasma, der 4. Aggregatzustand

4.1.1 Plasmazustand

In vielen Prozessen der Halbleiterherstellung wird ein Plasma eingesetzt, sei es beim Sputtern oder beim Trockenätzen. Ein wichtiger Punkt dabei ist, dass das Plasma nicht erhitzt ist. So können auch Wafer, die schon eine Metallisierungsschicht tragen, in Plasmaprozessen bearbeitet werden.

Plasma wird auch als der vierte Aggregatzustand bezeichnet. Ein Aggregatzustand ist ein qualitativer Zustand von Stoffen, der von der Temperatur und vom Druck abhängt. Den drei Aggregatzuständen fest, flüssig und gasförmig begegnet man auch im täglichen Leben. Ist die Temperatur niedrig, so ist jedes Atom in einem Stoff fest an einem Punkt angeordnet. Anziehungskräfte verhindern, dass sie sich bewegen. Beim absoluten Nullpunkt ($-273,15\text{ °C}$) gehen Stoffe auch keine Reaktion ein. Bei zunehmender Temperatur fangen die Teilchen an zu schwingen, und die Bindungen der Atome werden instabiler. Ist der Schmelzpunkt erreicht, geht ein Stoff vom ersten in den zweiten Aggregatzustand über: Eis (fest) geht über in Wasser (flüssig).

Die Anziehungskräfte in flüssigen Stoffen sind noch vorhanden, jedoch können sich die Teilchen verschieben und haben keine festen Plätze mehr wie im festen Zustand, die Teilchen passen sich beispielsweise einer vorgegebenen Form an. Wird die Temperatur noch weiter erhöht, brechen die Bindungen komplett auf, die Teilchen bewegen sich unabhängig voneinander. Am Siedepunkt geht ein Stoff vom zweiten in den dritten Aggregatzustand über: Wasser (flüssig) geht über in Wasserdampf (gasförmig).

Während das Volumen bei festen und flüssigen Stoffen konstant ist, nehmen gasförmige Stoffe den vorhandenen Platz vollständig ein, die Teilchen verteilen sich gleichmäßig im gesamten Raum.

Jeder Stoff hat einen ganz bestimmten Schmelz- und Siedepunkt. Silicium schmilzt bei 1414 °C und geht bei 2900 °C in den gasförmigen Zustand über. Führt man einem Stoff noch mehr Energie zu werden durch Zusammenstöße der Teilchen untereinander Elektronen aus der äußersten Elektronenschale herausgeschlagen. Es befinden sich nun freie Elektronen und positiv geladene Ionen im Raum: der Plasmazustand ist erreicht.

4.1.2 Plasmaerzeugung

Ein Plasma in der Halbleitertechnologie wird meist durch Hochfrequenzspannung erzeugt, als Gas dient z.B. Argon. Das Gas befindet sich in einem Hochfrequenzfeld zwischen zwei unterschiedlich geladenen Platten (Elektroden) und wird hier ionisiert. Dazu sind Elektronen notwendig, die aus den Argonatomen weitere Elektronen heraus schlagen. Diese ersten Elektronen können auf unterschiedliche Weise erzeugt werden:

- Elektronen werden von einer Glühkathode emittiert
- Durch sehr hohe Spannung werden Elektronen aus der negativen Elektrode herausgezogen
- In jedem Gas befinden sich durch Zusammenstöße der Teilchen kurzzeitig freie Elektronen

Da die Elektronen sehr viel leichter als die Ionen sind, werden sie sofort von der positiv geladenen Elektrode angezogen; die schweren Ionen bewegen sich langsam zur negativen Elektrode. Bevor sie diese jedoch erreichen werden die Elektroden umgepolt, die Elektronen werden zur anderen Elektrode gezogen und schlagen auf ihrer Flugbahn bei Zusammenstößen weitere Elektronen aus den Atomen heraus. Typische Frequenzen bei der Plasmaerzeugung sind 13,56 Megahertz und 2,45 Gigahertz, die Spannung an den Elektroden wird also 13,56 Millionen bzw. 2,45 Milliarden mal pro Sekunde umgepolt.

Die Elektronen befinden sich vornehmlich auf den Elektroden, während dazwischen die positiv geladenen Ionen, das Plasma, hin und her schwingen, da sie dem schnellen Spannungswechsel nicht folgen können.

Die Plasmaerzeugung findet unter Vakuum statt, das erzeugte Plasma ist dabei nicht erhitzt, was bei vielen Prozessen wichtig ist. Je nach Gas und Anlage in der das Plasma erzeugt wird, kann es in der Abscheidung, beim Sputtern, zum Ätzen oder auch

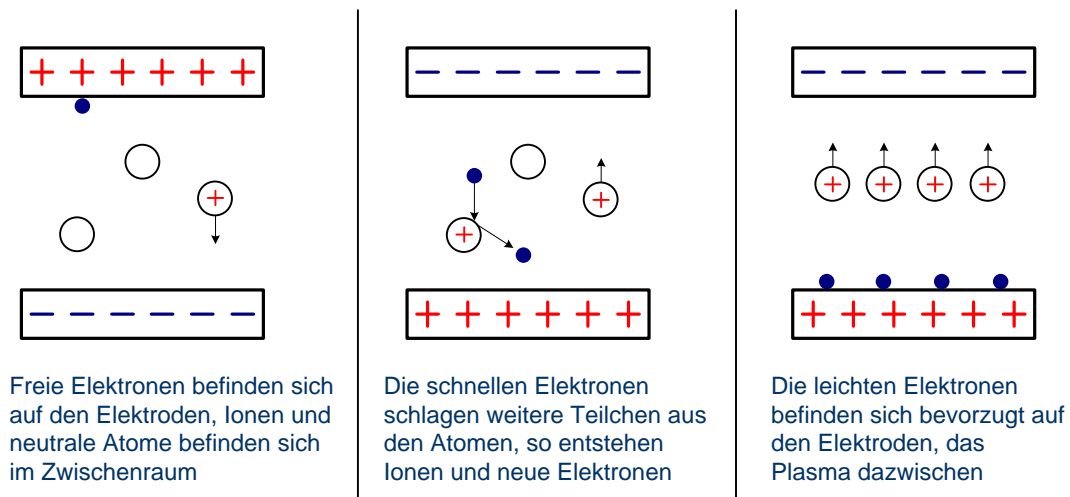


Abb. 4.1: Prinzip der Plasmaerzeugung

bei der Ionenimplantation verwendet werden. Durch die schnellen Schwingungen der positiven Ionen im Hochfrequenzfeld sind diese sehr energiereich. Es befinden sich dabei nicht nur positive Ionen und freie Elektronen im Plasma, da durch Zusammenstöße auch andere Teilchen erzeugt werden; der Zustand des Plasmas ändert sich laufend. Elektronen werden von den Ionen teilweise wieder eingefangen und erneut herausgeschlagen, diese zusätzlichen Teilchen spielen jedoch bei der weiteren Verwendung des Plasmas keine Rolle. Der Ionisierungsgrad beträgt je nach Teilchendichte im Plasma ($10^8 - 10^{12}$ Teilchen pro cm^3) 0,001-10 %, die Mehrzahl der Teilchen ist also ungeladen.

4.2 CVD-Verfahren

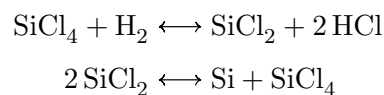
4.2.1 Siliciumgasphasenepitaxie

Epitaxie bedeutet „obenauf“ oder „zugeordnet“, und stellt einen Prozess dar, bei dem eine Schicht auf einer anderen Schicht erzeugt wird und deren Kristallstruktur übernimmt. Ist die abgeschiedene Schicht aus dem gleichen Material wie die Unterlage spricht man von Homoepitaxie, bei zwei unterschiedlichen Materialien von Heteroepitaxie. Der bedeutendste Fall bei der Homoepitaxie ist die Abscheidung von Silicium auf Silicium, bei der Heteroepitaxie wird meist eine Siliciumschicht auf einem Isolator wie z.B. Saphir erzeugt (Silicon On Insulator: SOI).

Die Homoepitaxie:

Je nach Prozess können die Wafer bereits vom Hersteller mit einer Epitaxieschicht angeliefert werden (z.B. bei CMOS), oder der Chiphersteller muss dies selbst durchführen (z.B. in der Bipolartechnik).

Als Gase zur Erzeugung der Schicht benutzt man reinen Wasserstoff in Verbindung mit Silan (SiH_4), Dichlorsilan (SiH_2Cl_2) oder Siliciumtetrachlorid (SiCl_4). Bei ca. 1000 °C spalten die Gase das Silicium ab, das sich auf der Waferoberfläche absetzt. Das Silicium übernimmt die Struktur des Substrats und wächst dabei, aus energetischen Gründen, Ebene für Ebene nacheinander auf. Damit das Silicium nicht polykristallin aufwächst, muss immer ein Mangel an Siliciumatomen vorherrschen, d.h. es steht immer etwas weniger Silicium zur Verfügung, als aufwachsen könnte. Beim Siliciumtetrachlorid verläuft die Reaktion in zwei Schritten:



Damit das Silicium auf dem Substrat die Struktur des Kristalls übernehmen kann, muss die Oberfläche absolut rein sein, dabei macht man sich die Gleichgewichtsreaktion zu Nutze. Beide Reaktionen können auch in die andere Richtung verlaufen, je nachdem wie das Verhältnis der eingesetzten Gase ist. Befindet sich nur wenig Wasserstoff in der Atmosphäre wird, wie beim Trichlorsilanprozess zur Reinigung des Siliciums, auf Grund der hohen Chlorkonzentration Silicium von der Waferoberfläche abgetragen. Erst mit zunehmender Konzentration von Wasserstoff wird ein Wachstum erzielt.

Die Aufwachsrate beträgt bei SiCl_4 ca. 1-2 μm pro Minute. Da das einkristalline Silicium nur auf der gereinigten Oberfläche aufwächst, kann man mit Oxid bestimmte Bereiche maskieren, auf denen dann polykristallines Silicium aufwächst. Dieses wird jedoch im Vergleich zu einkristallinem Silicium sehr leicht durch die rückwärts ablaufende Reaktion abgeätzt. Werden den Prozessgasen noch Diboran (B_2H_6) oder Phosphin (PH_3) hinzugefügt, kann man dotierte Schichten erzeugen, da sich die Dotiergase bei den hohen Temperaturen zersetzen und die Dotierstoffe im Kristallgitter eingebaut werden.

Der Prozess zur Herstellung von Homoepitaxieschichten wird im Vakuum durchge-

führt. Dabei wird die Prozesskammer zunächst auf 1200 °C aufgeheizt, damit sich das natürliche Oxid, das sich immer auf der Siliciumoberfläche bildet, verflüchtigt. Wie oben erwähnt, kommt es bei zu geringer Wasserstoffkonzentration zu einer Rückätzung der Wafer, das macht man sich vor dem eigentlichen Prozess zu Nutze, indem man auf diese Weise die Oberfläche der Wafer reinigt. Durch Veränderung der Gaskonzentrationen findet dann die Abscheidung in einer Epitaxieanlage statt.

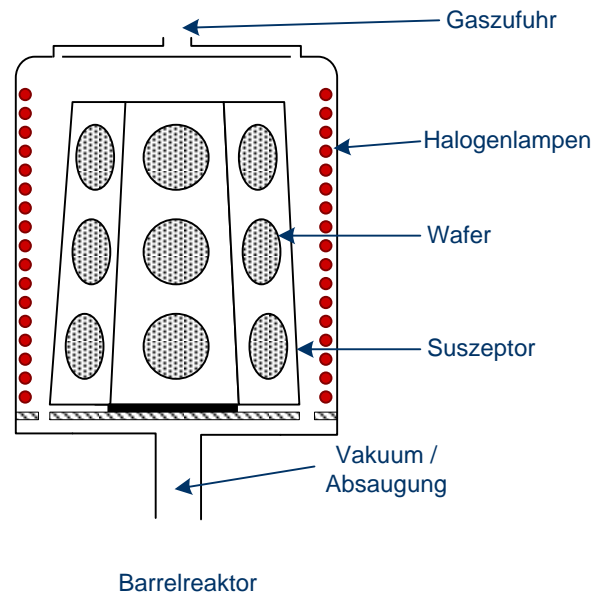


Abb. 4.2: Darstellung eines Barrelreaktors zur Epitaxieabscheidung

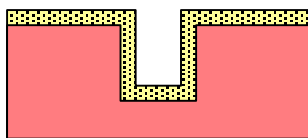
Durch die hohen Temperaturen kommt es jedoch zu einer Diffusion der Dotierstoffe im Substrat, bzw. zu einer Verunreinigung durch Dotierstoffe, die sich bei vorherigen Prozessen an den Reaktorwänden abgelagert haben. Bei der Verwendung von SiH_2Cl_2 oder SiH_4 sind solch hohe Temperaturen nicht nötig, weshalb man sich eher auf diese Gase beschränkt. Damit jedoch eine Scheibenreinigung durch Rückätzung stattfinden kann muss man zusätzlich als Gas HCl zugeben. Der Nachteil dieser Silane ist jedoch, dass sich bereits in der Atmosphäre, vor der Abscheidung auf dem Wafer, Keime bilden. Dadurch ist die erzielte Qualität der Schicht nicht so hoch.

4.2.2 CVD-Verfahren: Chemical Vapor Deposition

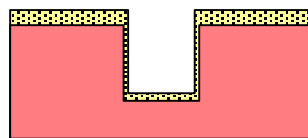
Oftmals werden in der Halbleitertechnologie Schichten benötigt, die nicht aus dem Siliciumsubstrat erzeugt werden können. Bei Siliciumnitrid und Siliciumoxinitrid muss

die Schicht z.B. durch thermische Zersetzung von Gasen erzeugt werden, die alle benötigten Materialien, wie Silicium, enthalten. Darauf basiert die Chemische Dampfab-scheidung, kurz CVD (engl.: Chemical Vapor Deposition). Der Wafer dient dabei nur als Grundfläche und reagiert nicht mit den Gasen. Dabei wird das CVD-Verfahren je nach Druck und Temperatur in drei Verfahren unterteilt, deren Schichten sich durch Dichte und Kantenbedeckung unterscheiden. Ist das Schichtwachstum an vertikalen Kanten genau so hoch wie auf horizontalen Flächen ist die Abscheidung konform.

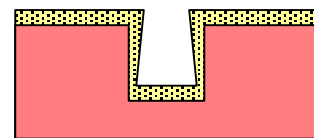
Die Konformität K ist der Quotient aus Schichtwachstum auf vertikalen Flächen R_v und dem Wachstum auf horizontalen Flächen R_h . $K = R_v/R_h$. Ist die Abscheidung nicht ideal konform, ist K deutlich kleiner als 1 (z.B. $R_v/R_h = 1/2 \Rightarrow K = 0,5$). Hohe Konformitäten lassen sich nur durch hohe Prozesstemperaturen erreichen.



Konforme Abscheidung
 $K = 1$



Nicht ideal konform
 $K = 0,5$

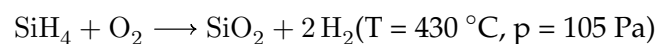


Ungleichmäßige
Beschichtung

4.2.3 APCVD: Atmospheric Pressure CVD

APCVD ist eine CVD-Abscheidung bei Normaldruck (atmosphärischer Druck) die zur Herstellung von dotierten und undotierten Oxiden eingesetzt wird. Das erzeugte Oxid hat eine geringe Dichte, die Kantenbedeckung ist auf Grund der niedrigen Temperatur mäßig. Durch verbessertes Equipment erlebt das APCVD-Verfahren aber eine Renaissance. Der hohe Waferdurchsatz ist ein großer Vorteil bei diesem CVD-Prozess.

Als Prozessgase dienen Silan SiH_4 (stark verdünnt mit Stickstoff N_2) und Sauerstoff O_2 die bei ca. 400°C thermisch zersetzt werden und miteinander reagieren.



Durch Zugabe von Ozon (O_3) kann eine bessere Konformität erzielt werden, da es die Beweglichkeit der sich anlagernden Teilchen erhöht. Das Oxid ist porös und elektrisch

instabil und kann durch einen Hochtemperaturschritt verdichtet werden. Damit sich keine Kanten bilden die Probleme bei der Erzeugung weiterer Schichten machen könnten, wird für Schichten, die als Zwischenoxid dienen, so genanntes Phosphorglas (PSG) eingesetzt. Dazu wird den zwei Gasen SiH_4 und O_2 zusätzlich noch Phosphin (PH_3) zugesetzt, so dass das entstandene Oxid 4-8 % Phosphor enthält. Ein hoher Phosphorgehalt erhöht die Fließeigenschaften deutlich, es kann sich allerdings Phosphorsäure bilden, die Aluminium (Leiterbahnen) korrodieren lässt.

Da Temperschritte aber auch vorangegangene Prozesse (z. B. Dotierungen) beeinflussen können, wird zum Verfließen des PSG nur ein kurzzeitiges Tempern mit einer extrem starken Argonlampe (mehrere hundert Kilowatt, <10 s, $T = 1100^\circ\text{C}$) an Stelle von langen Ofenprozessen angewandt. Analog zum Phosphorsilicatglas, kann gleichzeitig auch Bor hinzugefügt werden (Borphosphorsilicatglas, BPSG, 4 % B und 4 % P).

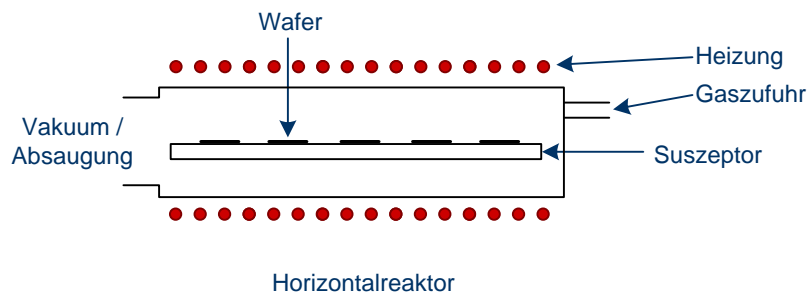


Abb. 4.3: Darstellung eines Horizontalreaktors zur APCVD-Abscheidung

4.2.4 LPCVD: Low Pressure CVD

Beim LPCVD-Verfahren findet der Prozess unter Vakuumatmosphäre statt, bei dem sich dünne Schichten wie Siliciumnitrid (Si_3N_4), Siliciumoxinitrid (SiON), SiO_2 und Wolfram (W) erzeugen lassen. LPCVD-Verfahren ermöglichen sehr gute Konformitäten von beinahe 1; Grund dafür ist der niedrige Druck von nur 10-100 Pa (normaler Luftdruck ca. 100.000 Pa), durch den die Teilchen keine Bewegung in einer Richtung erfahren, sondern sich selbst durch Zusammenstöße ausbreiten, somit werden die Kanten auf der Waferoberfläche von allen Gasteilchen gleichmäßig bedeckt. Die bis zu 900°C hohe Prozesstemperatur trägt zur hohen Konformität bei. Im Gegensatz zum APCVD-Verfahren ist die Dichte und Stabilität sehr hoch.

Die Reaktionen für Si_3N_4 , SiON , SiO_2 und Wolfram laufen nach den folgenden Reaktionsgleichungen ab:

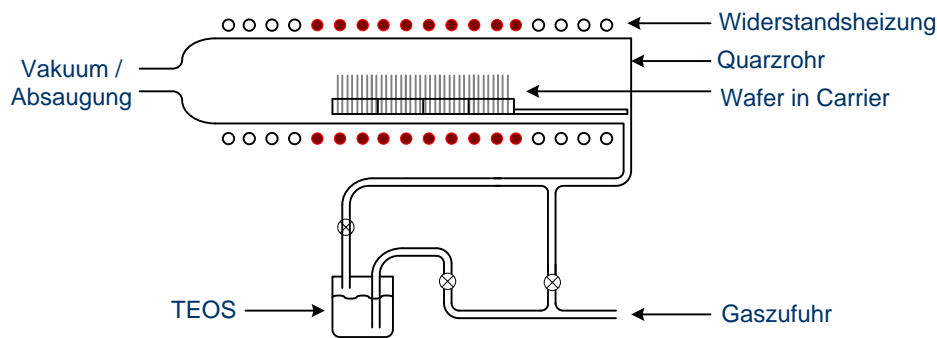


Abb. 4.4: Darstellung einer LPCVD-Abscheidungsanlage für TEOS-Schichten

- **Si₃N₄ (850 °C):** $4 \text{NH}_3 + 3 \text{SiH}_2\text{Cl}_2 \longrightarrow \text{Si}_3\text{N}_4 + 6 \text{HCl} + 6 \text{H}_2$
- **SiON (900 °C):** $\text{NH}_3 + \text{SiH}_2\text{Cl}_2 + \text{N}_2\text{O} \longrightarrow \text{Si}_3\text{N}_4 + \text{Nebenprodukte}$
- **SiO₂ (700 °C):** $\text{SiO}_4\text{C}_8\text{H}_{20} \longrightarrow \text{SiO}_2 + \text{Nebenprodukte}$
- **Wolfram (400 °C):** $\text{WF}_6 + 3 \text{H} \longrightarrow 2 \text{W} + 6 \text{HF}$

Im Gegensatz zu den gasförmigen Ausgangsstoffen bei Si₃N₄, SiON und Wolfram, wird zur Erzeugung von SiO₂ flüssiges Tetraethylorthosilicat (TEOS) verwendet; daneben gibt es noch weitere Flüssigquellen mit so unaussprechlichen Namen wie Diteriäbutylsilan (DTBS, SiH₂C₈H₂₀) oder Tetramethylcyclotetrasiloxan (Si₄O₄C₄H₁₆). Eine Wolframschicht lässt sich nur auf Silicium erzeugen. Somit muss, wenn kein Silicium als Grundfläche zur Verfügung steht, Silan hinzugefügt werden.

4.2.5 PECVD: Plasma Enhanced CVD

Das Plasma unterstützte CVD-Verfahren findet bei 250-350 °C statt. Da die Temperatur zu niedrig ist um die Gase zu zersetzen, wird es mit Hilfe einer Hochfrequenzspannung in den Plasmazustand versetzt. Im Plasmazustand ist es sehr energiereich und scheidet sich dann auf der Scheibe ab. Besonders Leiterbahnen aus Aluminium können keinen hohen Temperaturen ausgesetzt werden (Schmelzpunkt von Aluminium ca. 500 °C) weshalb das PECVD-Verfahren vor allem hier zur Abscheidung von SiO₂ und Si₃N₄ genutzt wird. Anstelle des SiH₂Cl₂ beim LPCVD-Verfahren wird Silan verwendet, da es sich bei den niedrigen Temperaturen leichter zersetzt. Die Konformität ist nicht so ideal wie beim LPCVD-Verfahren (ca. 0,6-0,8), jedoch ist die Abscheiderate mit 500 nm pro Minute deutlich höher.

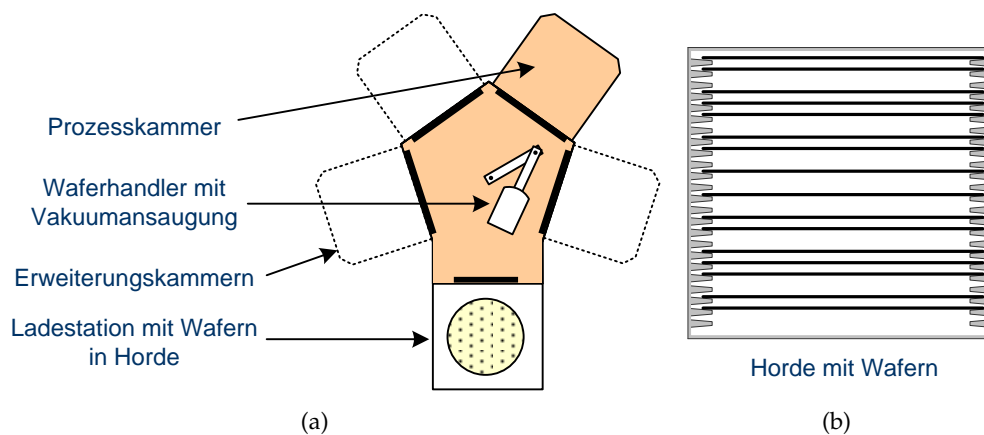


Abb. 4.5: (a) Darstellung einer PECVD-Abscheidungsanlage und (b) einer Horde mit Wafern.

4.2.6 ALD: Atomic Layer Deposition

Die Atomlagenabscheidung (Atomic Layer Deposition, kurz ALD) ist ein CVD-Abscheidungsverfahren zur Erzeugung dünner Schichten. Dabei handelt es sich um ein zyklisches Verfahren, bei dem mehrere Gase abwechselnd in die Prozesskammer eingeleitet werden.

Jedes Gas reagiert dabei in der Art, dass die jeweiligen Oberflächenatome abgesättigt werden, die Reaktion also selbstbegrenzend abläuft. Das jeweils andere Gas kann dann an dieser Oberfläche weiterreagieren. Alternierend zu den reaktiven Gasen wird die Kammer mit Inertgasen, wie Argon oder Stickstoff gespült. Ein einfacher ALD-Zyklus, der meist nur wenige Sekunden beträgt, kann daher wie folgt ablaufen:

1. selbst begrenzende Reaktion mit dem ersten Gas, welches die Oberflächenatome absättigt
2. Spülschritt mit einem Inertgas
3. selbst begrenzende Reaktion mit einem zweiten Gas, welches an der neuen Oberfläche reagiert
4. Spülschritt mit einem Inertgas

Ein konkretes Beispiel ist die Abscheidung einer Aluminiumoxidschicht (Al_2O_3), bei der als reaktive Gase Trimethylaluminium (TMA, $\text{C}_3\text{H}_9\text{Al}$) und Wasser (H_2O) verwendet werden.

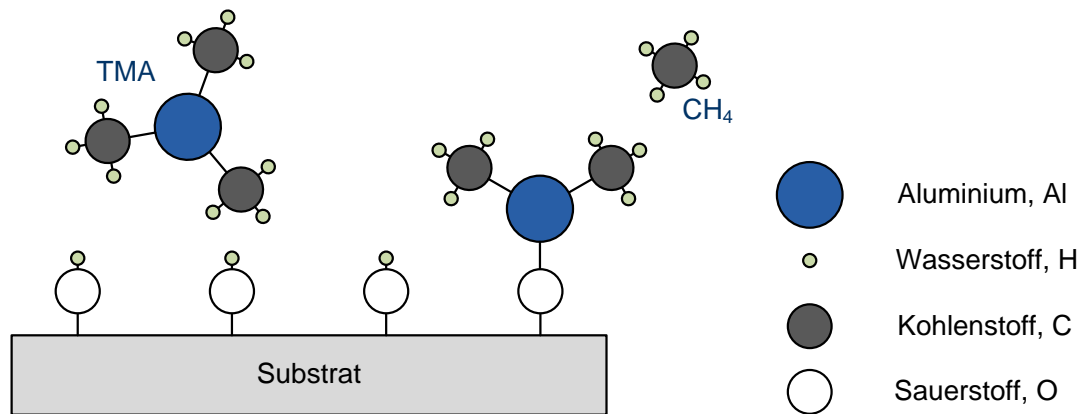


Abb. 4.6: Einlass von TMA in die Prozesskammer

Dabei entfernt eine Methylgruppe (CH_3) der Aluminiumverbindung im ersten Schritt oberflächennahe Wasserstoffatome unter Bildung von Methan (CH_4). Die verbleibenden Moleküle werden an den nun ungesättigten Sauerstoffatomen gebunden.

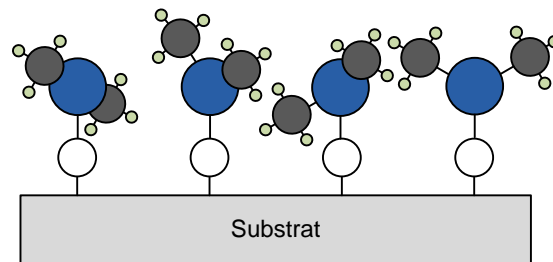


Abb. 4.7: Anlagerung von TMA an der Waferoberfläche

Sind diese Atome abgesättigt, kann keine weitere Anlagerung von TMA erfolgen.

Die Prozesskammer wird gespült und anschließend Wasserdampf in die Kammer eingeleitet. Durch je ein Wasserstoffatom der H_2O -Moleküle werden die CH_3 -Gruppen der zuvor abgeschiedenen Schicht unter Bildung von Methan entfernt.

Zurück bleiben OH^- -Ionen, welche sich nun mit dem Aluminium verbinden, so dass wiederum oberflächennahe Wasserstoffatome zur Verfügung stehen, die mit TMA reagieren können.

Die Atomlagenabscheidung bietet wesentliche Vorteile gegenüber anderen Abscheidungsverfahren, weshalb sie ein wichtiger Prozess zur Aufbringung verschiedener Schichten ist. Dabei können auch 3-dimensionale Strukturen (z. B. Gräben) sehr homogen

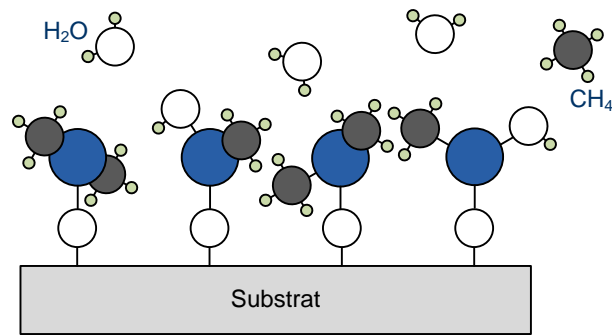


Abb. 4.8: Einlass von Wasserdampf in die Prozesskammer

beschichtet werden. Möglich sind sowohl isolierende als auch leitende Filme auf verschiedenen Substraten (Halbleiter, Polymere u. a.). Durch die Anzahl der Zyklen ist eine exakte Schichtdicke leicht zu realisieren. Da die reaktiven Gase nicht gleichzeitig in die Kammer geleitet werden, kann es nicht schon vor der eigentlichen Abscheidung zu Keimbildung kommen. Die Qualität der Schicht ist damit sehr hoch.

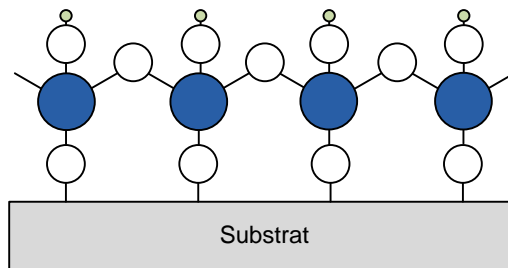


Abb. 4.9: Abgesättigte Oberfläche vor dem nächsten Zyklus

4.3 PVD-Verfahren

4.3.1 MBE: Molekularstrahlepitaxie

Das Aufwuchsverhalten der Molekularstrahlepitaxie (Molecular beam epitaxy, MBE) entspricht dem der Siliciumgasphasenepitaxie. Jedoch wird in diesem Fall das Silicium auf eine andere Weise auf dem Wafer abgeschieden.

Der Prozess findet unter Ultrahochvakuum statt (UHV, 10^{-8} Pa), der zu prozessierende Wafer wird mit der Oberseite nach unten an der Decke der Prozesskammer gehalten, und auf ca. 600-800 °C erhitzt damit sich das natürliche Oxid verflüchtigt. Silicium

wird dabei mit einem Elektronenstrahl in einem Evaporationstiegel verdampft, und scheidet sich dann auf dem Wafer ab. Dotierstoffe können in einer Effusorquelle verdampft werden und gelangen so mit dem Siliciumdampf auf den Wafer. Durch gezielte Temperaturregelung und mittels Blenden kann der Teilchenstrahl genau dosiert werden. Bei diesem Verfahren ist es außerdem möglich Schichten aus verschiedenen Stoffen übereinander wachsen zu lassen, die sonst auf Grund der unterschiedlichen Atomgrößen nicht möglich wären. So lassen sich Silicium und Germanium als Schicht erzeugen, die bei Hochfrequenzanwendungen in der Bipolartechnik von Nöten ist.

Jedoch ist der Aufwand bei dieser Epitaxie zur Erzeugung des UHV sehr hoch. Um diesen Druck, der weniger als dem billionsten Teil des normalen Luftdrucks entspricht, zu erreichen, müssen mehrere Pumpen kombiniert werden und die Prozesskammer lange abgepumpt werden. Dabei kann nur ein Wafer auf einmal prozessiert werden, das Wachstum beträgt nur ca. 1 μm pro Stunde.

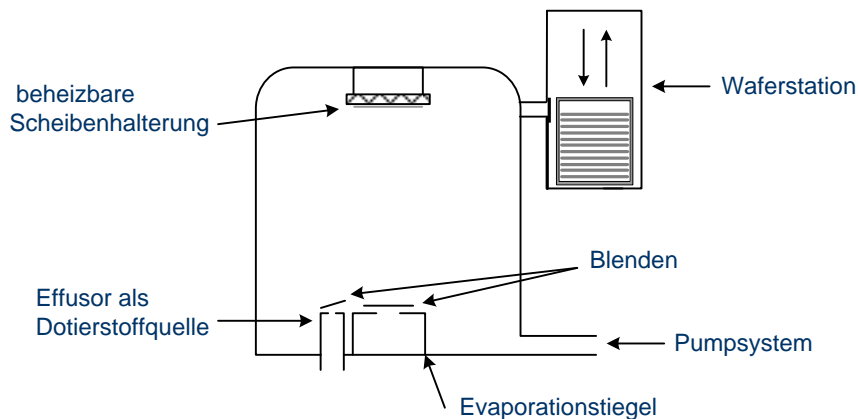


Abb. 4.10: Schematische Darstellung einer MBE-Anlage

4.3.2 Aufdampfen

Beim Aufdampfen können metallische Schichten, wie z.B. Aluminium, auf den Wafern erzeugt werden. Das Material befindet sich in einem Tiegel aus hochschmelzbarem Metall wie Tantal und wird dort erhitzt bis es in den gasförmigen Zustand übergeht. Der Metalldampf trifft senkrecht auf die Scheibe, so dass Kanten nicht gut bedeckt werden, die entstandene Schicht ist polykristallin. Alternativ zum Schmelzen im Tiegel kann das Metall auch mit einem Elektronenstrahl verdampft werden. Mit dieser Elektronenstrahlmethode ist es möglich eine, im Vergleich zur thermischen Aufdampfung,

sehr genaue Abscheiderate zu erzielen. Da die Kantenabdeckung nicht sehr gut ist, werden diese zwei Verfahren meist zur ganzflächigen Rückseitenbeschichtung für die spätere elektrische Kontaktierung verwendet.

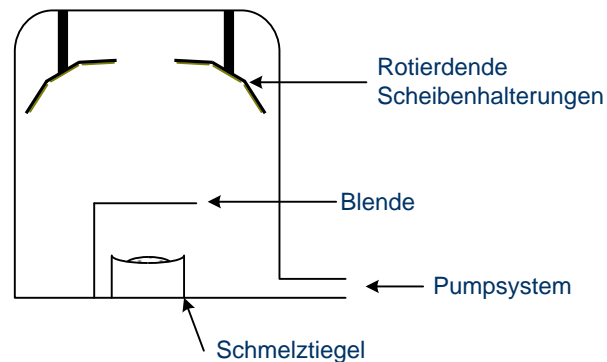
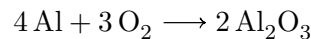
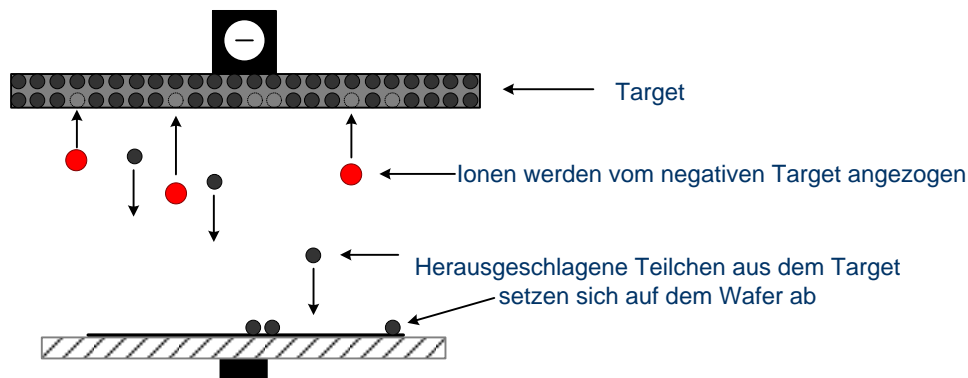


Abb. 4.11: Schematische Darstellung einer Aufdampfanlage

4.3.3 Sputtern: Kathodenzerstäubung

Bei der Kathodenzerstäubung werden Ionen auf ein Target beschleunigt und schlagen dort Atome oder Moleküle heraus, das Target besteht dabei aus dem Material der aufzubringenden Schicht. Die mittlere Weglänge dieser Teilchen ist wenige Millimeter lang, d.h. sie stoßen oft zusammen, dadurch werden auf der Scheibe auch vertikale Flächen gut bedeckt. Meist wird das Edelgas Argon durch Gasentladung ionisiert (freie Elektronen werden durch eine positive Spannung beschleunigt und schlagen Außenelektronen der Argonatome heraus). Die abgeschiedenen Teilchen bilden eine poröse Schicht, die durch Temperung verdichtet werden kann. Die Kathodenzerstäubung lässt sich in das passive (inerte) und das reaktive Sputtern unterteilen.

Beim passiven Sputtern wird nur das Material des Targets auf den Scheiben abgeschieden, entsprechend dem Material des Targets lassen sich hochreine Schichten erzeugen, da ein genaues Mischungsverhältnis der Stoffe im Target möglich ist. Beim reaktiven Sputtern wird dem Gas in der Prozesskammer ein Reaktionsgas (z.B. Sauerstoff O_2) hinzugefügt, dass sich mit dem zerstäubten Material des Targets verbindet und sich dann auf dem Wafer absetzt. Dadurch lassen sich aus einem Metalltarget (z.B. Aluminium Al) auch Isolierungsschichten herstellen:



Zur Erzeugung von metallischen Schichten wird das Gleichstrom-Sputtern (DC-Sputtern) verwendet. Dabei werden die Ionen mit bis zu 3 Kilovolt (kV) auf das Target beschleunigt und entladen sich dort. Da diese Ladungen immer abgeführt werden müssen, kann hier nur ein leitendes Material als Target dienen. Für isolierende Schichten muss das reaktive Sputtern eingesetzt werden. Will man eine isolierende Schicht direkt aus dem Target erzeugen verwendet man das Hochfrequenzsputtern (RF-Sputtern).

Beim RF-Sputtern liegt die Spannung an je einer Elektrode hinter dem Target (Kathode) und dem Wafer (Anode) an. Durch die hochfrequente Spannung werden die Elektronen bei der positiven Halbwelle am Target von diesem angezogen, wodurch sich das Target negativ auflädt. Durch dieses negative Target werden die Ionen angezogen und schlagen dort Teilchen heraus. Beim Magnetron-Sputtern befinden sich zusätzlich Magnete hinter dem Target, so dass Elektronen in Kreisbahnen abgelenkt werden und so häufiger Argonatome ionisieren können, was einen erheblichen Anstieg der Abscheiderate verursacht. Da die Anode mit der Prozesskammer verbunden ist, hat sie auf die Oberfläche bezogen im zeitlichen Mittel eine wesentlich geringere Potentialdifferenz gegenüber dem Plasma als die Kathode, weshalb die Ionen nur zum Target und nicht zum Wafer wandern.

Um die Kantenabdeckung zu erhöhen verwendet man das BIAS-Sputtern, bei dem am Substrat eine negative Spannung angelegt wird. Dadurch werden auch hier, wie am Target, Teilchen der erzeugten Schicht abgetragen und die Oberfläche eingeebnet. Jedoch muss darauf geachtet werden, dass kein Abtrag des Substrats erfolgt. Dieses so

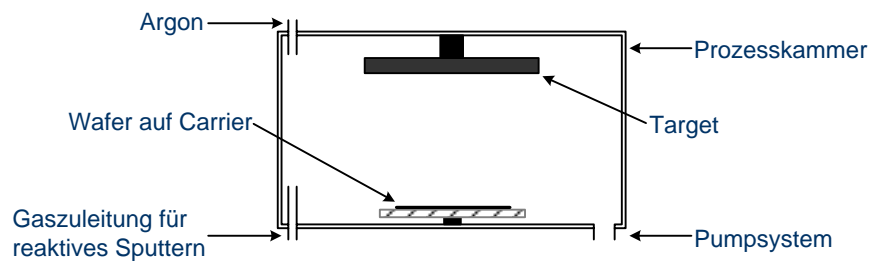


Abb. 4.12: Schematische Darstellung einer Sputteranlage

genannte Rückätzen ist auch das Prinzip der meisten Plasmaätzenanlagen.

Das Sputtern eignet sich somit zur Erzeugung von metallischen Schichten mit guter Konformität und sehr guter Reproduzierbarkeit. Der Aufwand ist gering, der Unterdruck mit 5 Pa recht einfach zu erzeugen.

5 Metallisierung

5.1 Anforderungen an die Metallisierung

Bei der Metallisierung werden Kontakte zu den dotierten Gebieten von Halbleiterbauelementen hergestellt, diese werden mit Leiterbahnen verbunden. Von dort werden die Anschlüsse zum Rand der einzelnen Chips geführt um die Verbindung vom Chip zum Gehäuse herzustellen oder zur Kontrolle der Chips mit Messsonden während der Fertigung.

Welche Voraussetzungen müssen die Metallisierungsebenen erfüllen um in der Halbleitertechnik eingesetzt werden zu können?

- gute Haftung auf Siliciumdioxid
- hohe Strombelastbarkeit, geringer elektrischer Widerstand
- geringer Kontaktwiderstand zwischen Metall und Halbleiter
- leichte Strukturierung in Trockenätzverfahren
- möglichst einfacher Prozess zum Aufbringen der leitfähigen Schicht
- geringe Korrosionsanfälligkeit für lange Lebensdauer der Chips
- gute Kontaktierbarkeit
- Möglichkeit der Mehrlagenverdrahtung zur Einsparung von Chipfläche
- hohe Reinheit des Materials

Da Aluminium viele dieser Anforderungen erfüllt hat es sich als bevorzugtes Metallisierungsmaterial durchgesetzt. Jedoch hat die Aluminiumtechnologie auch Nachteile; hier versucht man zukünftig mit Kupfer auszuweichen.

5.2 Aluminiumtechnologie

5.2.1 Aluminium und Aluminiumlegierungen

Wegen seiner Eigenschaften werden heutzutage Aluminium und Legierungen dieses Metalls sehr häufig an der Oberfläche der Chips zur Verdrahtung eingesetzt:

- gute Haftung auf SiO_2 und Zwischenoxiden wie BPSG und PSG
- gute Kontaktierbarkeit bei der Verdrahtung zum Gehäuse (z.B. mit Gold- und Aludrähten)
- niedriger spezifischer Widerstand von $3 \mu\Omega \cdot \text{cm}$
- sehr gut in Trockenätzverfahren strukturierbar

Aluminium erfüllt allerdings die Anforderung an elektrische Belastbarkeit und Korrosionsbeständigkeit nur teilweise. Metalle wie Silber oder Kupfer weisen hier bessere Eigenschaften auf, sind jedoch teuer und mit Trockenätzverfahren nur schwer strukturierbar.

5.2.2 Siliciumdiffusion

Bei der Verwendung von reinem Aluminium kann es zu einer Diffusion von Siliciumatomen in das Metall kommen. Der Halbleiter reagiert bereits bei 200-250 °C mit der Aluminiummetallisierung und hinterlässt durch den Materialschwund, verursacht durch die ausdiffundierten Atome, Gruben an der Kontaktfläche zwischen Silicium und Aluminium. Das Aluminium füllt diese Gruben auf. Dadurch entstehen „Spikes“ die unter Umständen zu einer Kurzschlussbildung führen, wenn sie durch die dotierten Gebiete bis in den Siliciumkristall hineinragen.

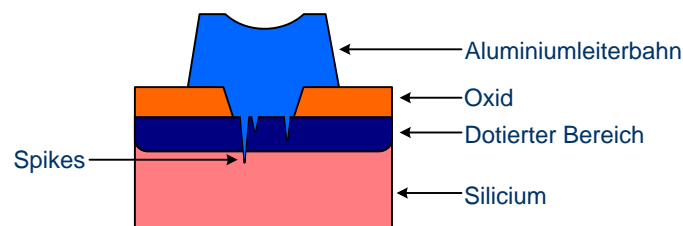


Abb. 5.1: Spikebildung

Die Größe dieser Spikes ist von der Temperatur abhängig mit der das Aluminium auf dem Silicium aufgebracht wird. Um diese Spikes zu verhindern gibt es mehrere Möglichkeiten. An der Stelle des Kontaktlochs kann eine tiefe Ionenimplantation, die Kontaktimplantation, eingebracht werden. Somit ragen die Spikes nicht bis in das Substrat hinein.

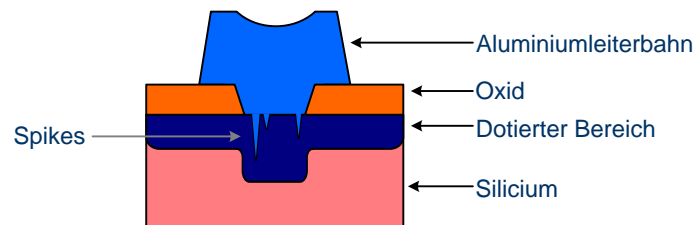


Abb. 5.2: Kontaktimplantation

Der Nachteil dabei ist jedoch, dass ein weiterer Prozessschritt eingeführt werden muss, und sich die elektrischen Eigenschaften durch die Vergrößerung des dotierten Gebietes ändern.

Anstelle des reinen Aluminiums kann auch eine Aluminium-Silicium-Legierung verwendet werden die ca. 1-2 % Siliciumanteil enthält. Das Aluminium ist nun bereits mit Silicium versetzt und es diffundieren keine Siliciumatome mehr aus dem Wafer in das Aluminium. Bei sehr kleinen Kontaktlöchern kann jedoch Silicium an der Kontaktfläche ausfallen, was in einem vergrößerten Kontaktwiderstand resultiert.

Für hochwertige Kontakte ist eine Trennung zwischen Aluminium und Silicium erforderlich. Dazu bringt man auf dem Silicium eine Barriere aus verschiedenen Stoffen, wie z.B. Titan, Titanitrid oder Wolfram auf. Damit eine Erhöhung des Kontaktwiderstands an der Grenzschicht zwischen Titan und Silicium unterbunden wird, muss hier noch eine Kontaktschicht aus Titansilicid aufgebracht werden.

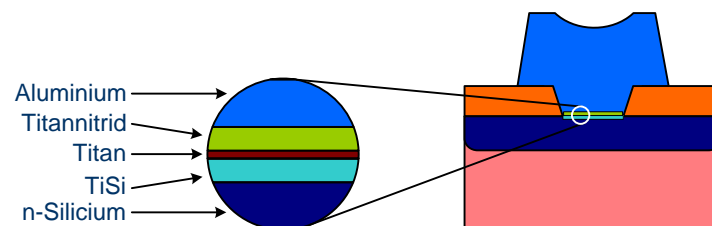


Abb. 5.3: Barrierschicht zwischen Aluminium und Substrat

5.2.3 Elektromigration

Bei hohen Stromdichten (Stromfluss pro Fläche) tritt eine Reibung zwischen den Elektronen und den festen Atomrümpfen auf, die Atome werden von ihren Plätzen weg bewegt. Besonders an Stellen mit geringem Leiterbahnquerschnitt ist die Stromdichte erhöht, durch die Verschiebung der Atome nimmt der Querschnitt ab, die Stromdichte steigt weiter an. Insbesondere an Kanten über die die Leiterbahnen verlaufen sind solche Engstellen zu finden. Im Extremfall reißen die Aluminiumleiterbahnen durch den Materialtransport ab.

5.2.4 Hillockwachstum

Durch die Elektromigration wird Material verschoben und an Stellen geringer Stromdichte angehäuft. Durch diese so genannten Hillocks (Hügel) können darüber liegende Schichten durchbrochen werden, was zu einem Kurzschluss mit einer anderen Metallisierungsebene führen kann, außerdem kann Feuchtigkeit durch die Risse eindringen und Korrosion verursachen. Hillocks entstehen aber auch auf Grund unterschiedlicher Ausdehnungskoeffizienten der Materialien. Die Stoffe dehnen sich bei Temperaturänderungen unterschiedlich aus, und es entstehen Spannungen zwischen den Schichten. Mit Ausgleichsschichten, die einen „mittleren“ Ausdehnungskoeffizienten haben kann dieses Problem gelöst werden (z.B. Titan, Titannitrid).

Weitere negative Effekte die bei der Metallisierung auftreten können:

- **Streubelichtung:** das Metall kann bei unebenen Oberflächen einfallende Lichtstrahlen bei der Belichtung schräg reflektieren, so dass auch Bereiche neben der eigentlichen Belichtung belichtet werden. Mit Hilfe einer Antireflexschicht aus amorphem Silicium wird die Reflexion verhindert
- **Schlechte Kantenabdeckung:** an Kanten kann es zu vermehrtem, in Ecken zu vermindertem Aufwachsen der Schicht kommen. Die Kanten müssen deshalb verrundet werden:

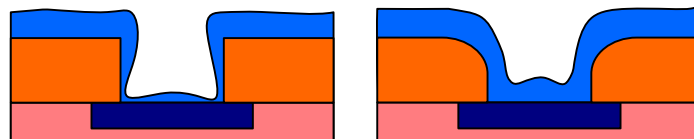


Abb. 5.4: Kantenverrundung zur konformen Aluminiumabscheidung

Das Design der Leiterbahnen muss also exakt geplant werden um diesen Fällen vorzubeugen. Durch einen geringen Zusatz an Kupfer kann die Lebensdauer der Aluminiumleiterbahnen stark erhöht werden, allerdings ist die Strukturierbarkeit mit Kupfer wesentlich komplizierter. Zum Schutz vor Korrosion werden die Oberflächen mit Siliciumdioxid, Siliciumtetranitrid oder Siliciumnitrid passiviert. Das Material der Gehäuse für die Chips besteht aus einer Art Keramik, da Kunststoff keinen ausreichenden Schutz vor Korrosion an den Stellen der Verdrahtung nach außen bietet. Verfahren wie Metallisierungsschichten auf dem Wafer aufgebracht werden, sind im Kapitel Abscheidung näher beschrieben.

5.3 Kupfertechnologie

5.3.1 Kupfertechnologie

Kupfer bietet gegenüber Aluminium wesentliche Vorteile, und löst das Leichtmetall bei kleinen Strukturen ab. So bietet das Metall einen wesentlich geringeren Widerstand als Aluminium, bei dem gut 50 % der zugeführten Leistung in Wärme umgewandelt werden. Da die Integration von Bauteilen ständig steigt, müssen die Leiterbahnen kleiner werden. Aluminium erfüllt, auch mit Kupferanteil, ab Strukturen unter 250 μm kaum mehr die benötigten elektrischen Anforderungen. Auch die Elektromigration ist bei Kupfer wesentlich weniger stark ausgeprägt. Ein Umstieg auf Kupfer ist somit unausweichlich, wenn die Strukturgrößen weiter abnehmen.

Kupfer hat jedoch die negative Eigenschaft, dass es nahezu alles mit dem es in Kontakt kommt kontaminiert. So müssen die Bereiche und Anlagen in der Fertigung in denen mit Kupfer gearbeitet wird von den anderen strikt abgetrennt werden. Zudem korrodiert Kupfer sehr leicht und muss deshalb, wie auch Aluminium, mit einer Passivierungsschicht versiegelt werden. Während bei Aluminium eine Durchkontaktierung der Ebenen mit Wolfram geschieht, wird bei Kupfer das Metall selbst dazu verwendet (lediglich die erste Schicht beim Kontakt der dotierten Siliciumgebiete erfordert eine Trennung mit Wolfram). So entfallen nicht nur die negativen thermoelektrischen Effekte die an den Übergängen zwischen verschiedenen Metallschichten entstehen, sondern auch die zusätzlichen Arbeitsschritte zum Aufbringen mehrerer Schichten. Jedoch lässt sich Kupfer im Vergleich zu Aluminium nur schlecht in Ätzverfahren strukturieren.

Das „gewöhnliche“ subtraktive Verfahren zur Strukturierung einer Schicht, wie es

auch bei Aluminium angewandt wird, sieht vereinfacht so aus:

- zu strukturierende Schicht aufbringen
- Fotolack aufbringen, belichten und entwickeln
- Lackmaske in einem Ätzschritt in die darunterliegende Schicht übertragen
- Lack entfernen
- Passivierungsschicht aufbringen

Bei der Metallisierung mit Kupfer bedient man sich eines additiven Verfahrens, dem so genannten Damascene-Prozess.

5.3.2 Damascene-Verfahren

Beim Damascene-Verfahren werden in bestehende Zwischenschichten, die zur Isolierung bzw. Passivierung dienen, die Kontaktlöcher (VIAs = Vertical Interconnect Access) der einzelnen Metallisierungsebenen geätzt, sowie die Gräben (Trenches), in denen die Kupferleiterbahnen verlaufen werden. In die Öffnungen kann nun mittels CVD-, PVD- + Reflow- sowie elektrochemischen (galvanischen) Verfahren Kupfer abgeschieden werden. Anschließend wird das Kupfer in einem CMP-Prozess planarisiert, bis die Oberfläche eingeebnet ist.

Man unterscheidet dabei noch zwischen Single- und Dual-Damascene-Prozess und bei letzterem noch zwischen VFTL (VIA First Trench Last: VIA zuerst, Graben zuletzt) und TFVL (Trench First VIA Last: Trench zuerst, VIA zuletzt). Im Folgenden werden die zwei Dual-Damascene-Verfahren näher erläutert.

Dual-Damascene: Trench First VIA Last:

Auf dem Wafer (hier auf einer bereits bestehenden Kupferebene) werden verschiedene Materialien aufgebracht, die als Isolations-, Passivierungs- oder Schutzschicht dienen. Als Ätzstopp und Schutz vor Ätzgasen kann Siliciumnitrid (SiN) zum Einsatz kommen. Als dielektrische Zwischenschicht (interlayer dielectric, ILD) kommen Materialien zum Einsatz, die einen geringen k-Wert haben, wie Siliciumdioxid SiO₂. Darüber wird eine Lackmaske strukturiert. Details zu den Schichten werden im nachfolgenden Abschnitt Low-k-Technologie erläutert.

1. Der Wafer wird mit Resist beschichtet, der in einem Lithografieprozess strukturiert wird (Abb. 5.5).

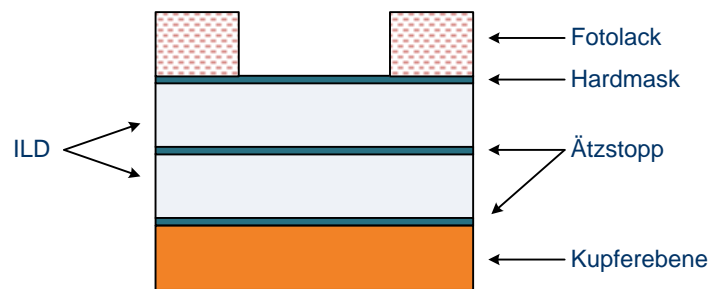


Abb. 5.5: Prinzipieller Damascene-Schichtaufbau vor der Strukturierung

2. In einem anisotropen Ätzprozess wird durch die Hartmaske (SiN) und die ILD-Schicht bis zur ersten Ätzstoppschicht (ebenfalls SiN) geätzt. Der Fotoresist wird entfernt, und der Graben für die Leiterbahnen ist fertig strukturiert (Abb. 5.6(a)).

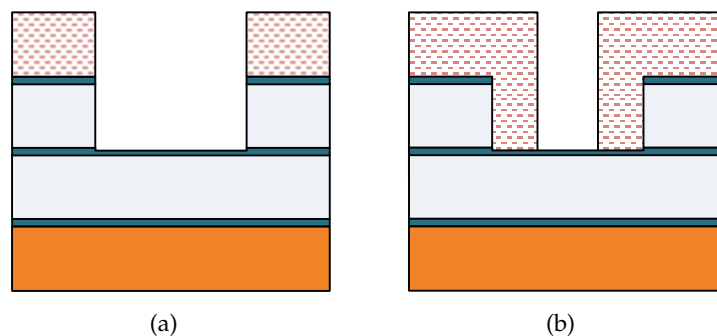


Abb. 5.6: Grabenätzung und VIA-Lackmaske

Die Hartmaske an der Oberfläche wird benötigt, um die ILD-Schicht während des Strippens (Lackentfernen) vor dem Plasma zu schützen. Dies ist notwendig, da die ILD-Schicht chemisch ähnlich aufgebaut ist wie der Fotolack, und so durch die gleichen Prozessgase angegriffen werden kann. Zusätzlich dient die Hartmaske als CMP-Stopp beim abschließenden Einebnen des Kupfers.

3. Als nächstes wird erneut Lack aufgebracht und strukturiert (Abb. 5.6(b)).

4. Anschließend werden in einem anisotropen Ätzprozess die Kontaktlöcher (VIAs) bis zum untersten Ätzstopp geöffnet. In einem niederenergetischen Ätzprozess wird dann die Ätzstoppschicht geöffnet, um kein darunterliegendes Kupfer herauszuschlagen, welches sonst in die ILD-Schicht eindiffundieren kann, danach wird der Fotolack entfernt. Als nächstes wird eine dünne Schicht Tantal abgeschieden, die verhindert, dass das Kupfer, welches in einem späteren Prozess aufgebracht wird, in die ILD-Schicht eindringt (Abb. 5.7(a)).

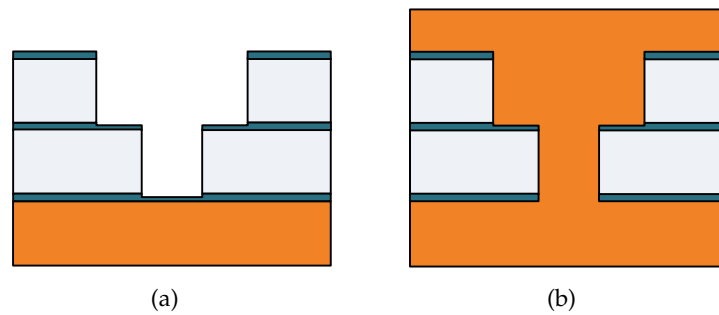


Abb. 5.7: Finale Struktur nach dem Ätzen und der Kupferabscheidung

5. Eine dünne Kupferkeimschicht wird aufgebracht und die Strukturen in einem galvanischen Verfahren mit Kupfer aufgefüllt (Abb. 5.7(b)).

6. Das Kupfer wird abschließend mittels CMP planarisiert (Abb. 5.8).

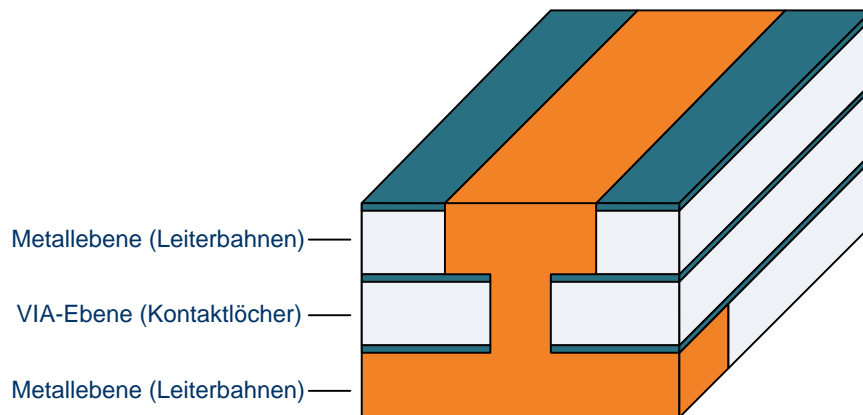


Abb. 5.8: Damascene-Struktur nach Metallisierung und Planarisierung

Der größte Nachteil beim TFVL-Verfahren ist die dicke Lackschicht die nach dem Ätzen der Gräben aufgebracht werden muss (3). Die winzigen VIAs in einer so dicken Lackschicht korrekt herzustellen ist extrem schwer. Das TFVL-Verfahren wird daher nur für die obersten Metallisierungsschichten verwendet, bei denen die Abmessungen der Strukturen nicht so kritisch sind, wie in den untersten Lagen.

Dual-Damascene: VIA First Trench Last:

Der VFTL-Prozess ähnelt dem TFVL-Prozess mit dem Unterschied, dass zuerst die Kontaktlöcher und danach die Gräben strukturiert werden.

1. Zunächst wird eine Lackmaske für die VIAs strukturiert und die Kontaktlöcher in einem anisotropen Ätzschritt bis zum untersten Ätzstopp geöffnet. Wichtig ist, dass

der Ätzstopp nicht durchgeätzt wird, um kein darunterbefindliches Kupfer herauszuschlagen, welches sonst in die ILD-Schicht eindiffundiert (Abb. 5.9).

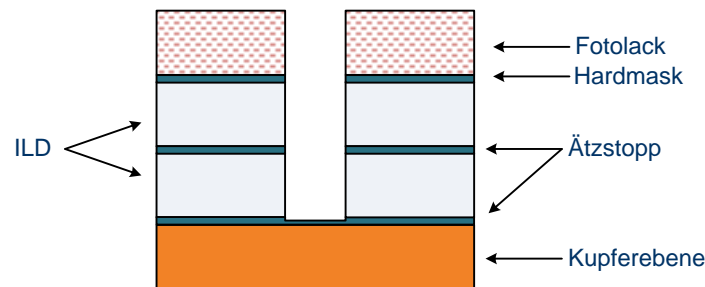


Abb. 5.9: Damascene-Schichtaufbau nach dem Kontaktlochätzen bei VFTL

2. Anschließend wird der Fotolack entfernt, und für die Gräben eine neue Lackmaske strukturiert; hierbei werden auch die bereits geöffneten VIAs mit Lack gefüllt (Abb. 5.10(a)).

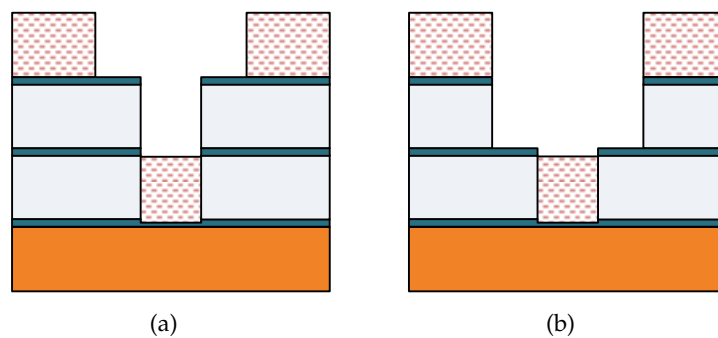


Abb. 5.10: Fototechnik und Ätzung der Gräben

3. Der unterste Ätzstopp wird durch den aufgetragenen Lack im VIA bei der anschließenden Grabenätzung vor den Ätzgasen geschützt. Danach wird analog zum TFVL-Prozess der unterste Ätzstopp geöffnet, eine Tantalbarriere und die Kupferkeimschicht aufgebracht (5.10(b)).

4. Als letztes folgt die Kupferabscheidung und die Planarisierung mittels chemisch mechanischem Polieren (Abb. 5.11).

Beim Single-Damascene-Verfahren werden die Ebenen für VIAs und Gräben einzeln abgeschieden und strukturiert. Dadurch sind zwei Kupferprozesse (Abscheidung, Planarisierung) notwendig.

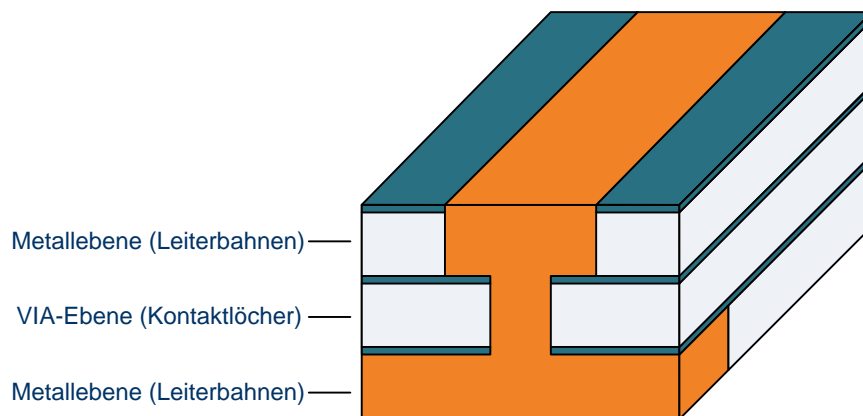


Abb. 5.11: Damascene-Struktur nach Metallisierung und Planarisierung

5.3.3 Low-k-Technologie

Durch die stetige Verkleinerung von Strukturen auf Mikrochips, um einerseits die Leistungsaufnahme zu verringern und andererseits maximale Schaltgeschwindigkeiten zu erzielen, rücken die Leiterbahnen zur Verdrahtung der einzelnen Bauelemente sowohl in vertikaler als auch lateraler Richtung immer näher zusammen. Um die Leiterbahnen von einander zu isolieren, müssen Schichten wie bspw. Siliciumdioxid SiO_2 als ILD aufgebracht werden.

Dort, wo Leiterbahnen parallel verlaufen oder sich auf übereinanderliegenden Metallisierungsebenen kreuzen, entstehen parasitäre Kapazitäten (Kondensatoren). Die Leiterbahnen bilden die leitenden Elektroden, das dazwischen liegende SiO_2 das Dielektrikum.

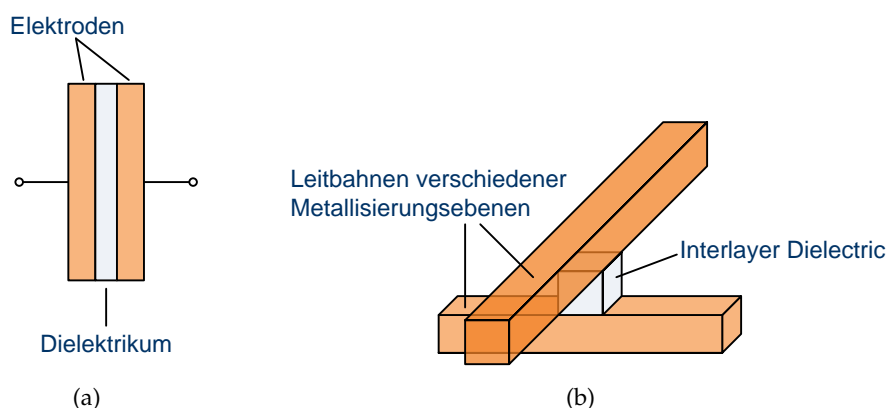


Abb. 5.12: (a) Klassischer Plattenkondensator und (b) Kapazität zwischen Leiterbahnen

Die Kapazität C eines Kondensators berechnet sich nach:

$$C = \frac{\epsilon_0 \epsilon_r A}{d}$$

Dabei steht d für den Abstand der Elektroden, A für die Fläche der Elektroden, also der sich überschneidenden Leiterbahnen, ϵ_0 für die absolute Dielektrizitätskonstante des Vakuums und ϵ_r (im Englischen häufig κ (Kappa) bzw. vereinfacht k) für die Dielektrizitätszahl des Isolators (hier SiO_2).

Die Größe der parasitären Kapazität beeinflusst nun die elektrischen Eigenschaften wie die maximale Schaltgeschwindigkeit oder den Stromverbrauch des Chips, weshalb versucht wird C möglichst klein zu halten. Dies ist theoretisch möglich durch eine Erniedrigung von ϵ_0 , ϵ_r und A oder durch eine Erhöhung von d . Da d wie zu Beginn erläutert immer kleiner wird, A durch die elektrischen Anforderungen vorgegeben ist und ϵ_0 eine physikalische Konstante ist, ergibt sich, dass die Kapazität eines Kondensators im Wesentlichen nur durch eine Verringerung von ϵ_r gesenkt werden kann.

Man benötigt also Dielektrika mit niedrigem ϵ_r : *low-k*.

Das klassische Dielektrikum, SiO_2 , hat eine Dielektrizitätszahl von ca. 4. Low-k beschreibt nun Materialien mit einem Wert $\epsilon_r < 4$, darüber hinaus werden Ultra-low-k-Materialien (ULK) zum Einsatz kommen, die ein $\epsilon_r < 2,4$ haben. Die Dielektrizitätszahl gibt die Polarisierung (Verschiebung von Ladungen innerhalb eines Isolators) im Dielektrikum an, und ist der Faktor, um den die Ladung einer Kapazität im Vergleich zu leerem Raum ansteigt oder um den das elektrische Feld im Kondensator abgeschwächt wird.

Um die Dielektrizitätszahl zu erniedrigen, gibt es im Grunde zwei Ansätze:

- die Polarisierbarkeit von Bindungen im Dielektrikum erniedrigen
- die Anzahl an Bindungen durch poröse Materialien reduzieren

Die Polarisierbarkeit kann durch Stoffe mit wenig polaren Gruppen gesenkt werden, möglich sind fluorierte (FSG, ϵ_r ca. 3,6) oder organische (OSG) Siliciumoxide. Dies allein ist jedoch bei den immer kleiner werdenden Strukturgrößen nicht mehr ausreichend, weshalb der Trend zu porösen Schichten geht. Durch die Porosität befindet sich dann innerhalb des ILD „leerer“ Raum, der im Falle von Luft eine Dielektrizitätszahl

von ca. 1 aufweist (besser ist nur Vakuum mit $\epsilon_r = 1$). Dadurch erniedrigt sich ϵ_r für die gesamte Schicht. Die Poren können erzeugt werden, indem das ILD-Material mit Polymeren versetzt wird, die in einem Tempersschritt aus der Schicht getrieben werden. Im Falle von SiO_2 müssten ca. 50 % Poren im Material eingebracht werden, damit man einen k-Wert von 2,0 erreicht. Geht man bereits von einem organischen Material aus, dessen k-Wert ohne Poren 2,5 beträgt, so müssten nur ca. 22 % Poren eingebracht werden.

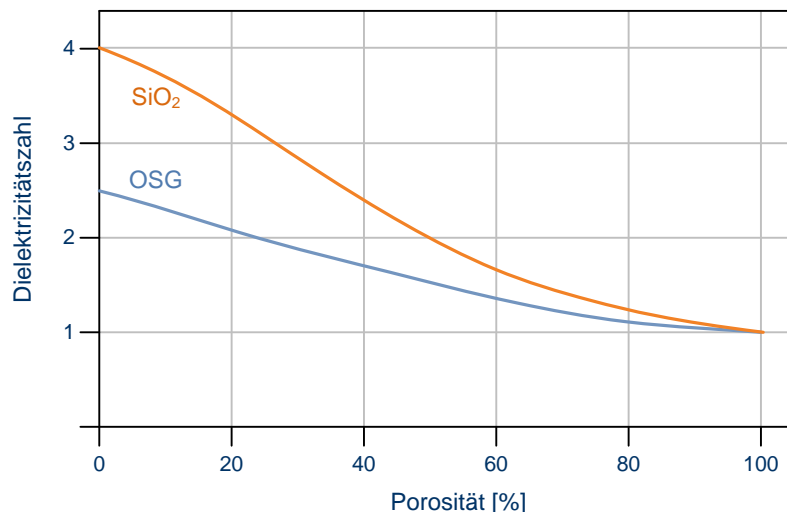


Abb. 5.13: k-Wert in Abhängigkeit von der Porosität

Jedoch ergeben sich einige Probleme die bewältigt werden müssen, um diese neuen Materialien in der Fertigung von Halbleitern einsetzen zu können.

Durch Poren im Material verringert sich dessen Dichte, was in geringerer mechanischer Stabilität resultiert. Zudem können Prozessgase oder Kupfer aus den Leiterbahnen leichter in die Schicht eindringen und diese schädigen, wodurch die Dielektrizitätszahl wieder ansteigt. Um dem entgegenzuwirken müssen die Poren möglichst gleichmäßig verteilt sein und dürfen nicht miteinander verbunden sein. Um eine Diffusion von Kupfer in das ILD zu verhindern, muss in einem zusätzlichen Prozessschritt eine dünne Diffusionsbarriere aufgebracht werden, jedoch muss darauf geachtet werden, dass dieses Material die Dielektrizitätszahl nicht erhöht.

Ebenso wie der in der Halbleiterfertigung eingesetzte Fotolack, bestehen auch die organischen Siliciumoxide aus Kohlenwasserstoffgruppen (CH). Wird nun der Lack nach der Strukturierung des Dielektrikums in einem Sauerstoffplasma entfernt, greift das

Plasma auch das Dielektrikum an. Auch hier müssen zusätzliche Schutzschichten (s. SiN im Abschnitt Damascene-Verfahren) aufgebracht werden die verhindern, dass die Isolationsschicht geschädigt wird.

Chemische Formel	k-Wert
SiO ₂	4,0
SiO _{1,5} CH ₃	3,0
SiO(CH ₃) ₂	2,7
SiO _{0,5} (CH ₃) ₃	2,55

Tab. 5.1: Übersicht verschiedener organischer Siliciumoxide

5.4 Der Metall-Halbleiter-Kontakt

5.4.1 Der Metall-Halbleiter-Kontakt

Nach der Herstellung der Transistoren im Siliciumsubstrat müssen diese mittels elektrischer Kontakte miteinander verbunden werden, um eine funktionstüchtige Schaltung zu erhalten. Dabei wird zum einen die Gateelektrode zur Steuerung des Transistors kontaktiert, zum anderen müssen die dotierten Source- und Draingebiete, über die der Stromfluss erfolgt, angeschlossen werden. Hier ergeben sich Probleme an den Kontaktflächen beim Übergang von Metallisierung zu Silicium, da je nach Dotierungstyp von Source und Drain Elektronenmangel (p-dotiert) oder Elektronenüberschuss (n-dotiert) vorherrscht.

Dabei spielt das sogenannte Fermi-niveau eine wichtige Rolle. Das Fermi-niveau ist das Energieniveau, bis zu dem sich am Temperaturnullpunkt (-273,15 °C) noch Elektronen befinden. In Leitern befinden sich Elektronen im Valenzband und im energetisch höheren Leitungsband, folglich ist das Fermi-niveau auf Höhe des Leitungsbandes. Zur Veranschaulichung: Ein See hat eine Wasseroberfläche, darunter befinden sich Wassermoleküle, welche die Elektronen darstellen. Diese reichen bis an die Oberfläche, das Fermi-niveau.

In dotierten Halbleitern befinden sich Fremdatome als Donatoren oder Akzeptoren im Kristallgitter. In einem p-dotierten Halbleiter ist das Fermi-niveau, auf Grund der Fähigkeit der Akzeptoratome Elektronen aus dem Valenzband aufzunehmen, knapp

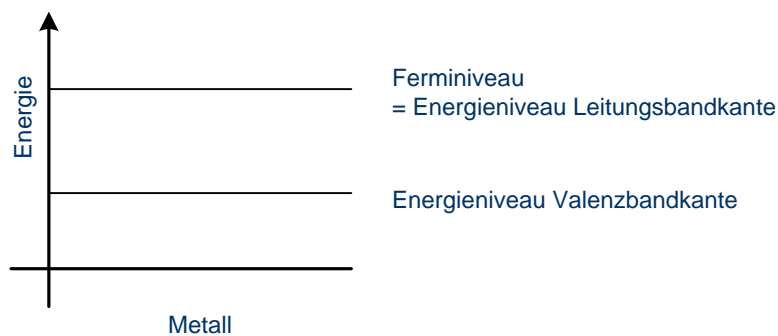


Abb. 5.14: Fermi-niveau in Metallen

über der Valenzbandkante. Dementsprechend liegt das Fermi-niveau in n-dotierten Halbleitern in der Nähe der Leitungsbandkante.

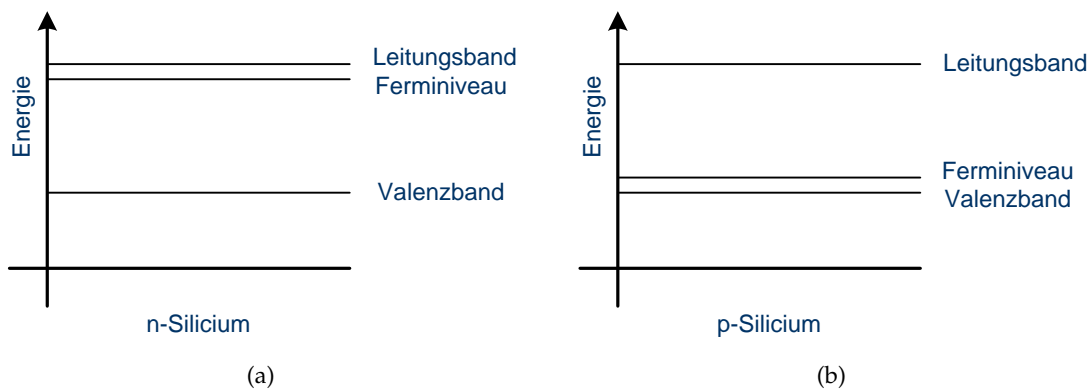


Abb. 5.15: Fermi-niveau in dotierten Halbleitern

Bringt man das Metall und den Halbleiter in Kontakt müssen sich die Fermi-niveaus angleichen, im Bereich der Grenzfläche ist das Fermi-niveau also konstant.

Da das Leitungsband im Halbleiter energetisch höher liegt als das Fermi-niveau, fließen beim Kontakt Elektronen vom Silicium in das Metall, da diese den energetisch niedrigsten Zustand einnehmen. Dadurch verringert sich gleichzeitig die Aufenthaltswahrscheinlichkeit von Elektronen im Leitungsband des Halbleiters, und das Fermi-niveau, welches den höchsten noch mit Elektronen besetzten Energiezustand beschreibt, entfernt sich. Durch die abgeflossenen negativen Ladungsträger bleiben positive Donatoren (z.B. Phosphorionen) zurück und es entsteht eine Raumladungszone. Die Verbiegung des Leitungsbandes veranschaulicht die Spannungsbarriere (Schottky-Barriere)

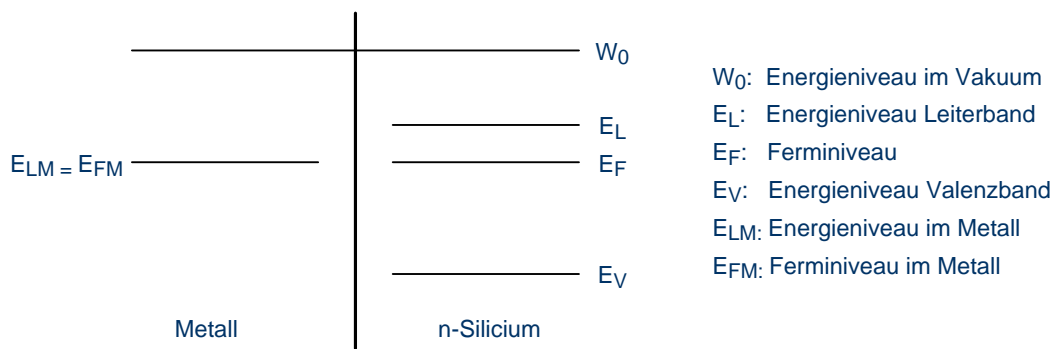


Abb. 5.16: Bändermodell vor dem Kontakt von Metall und n-Silicium

welche die verbliebenen Elektronen im n-Leiter überwinden müssen, um in das Metall zu fließen.

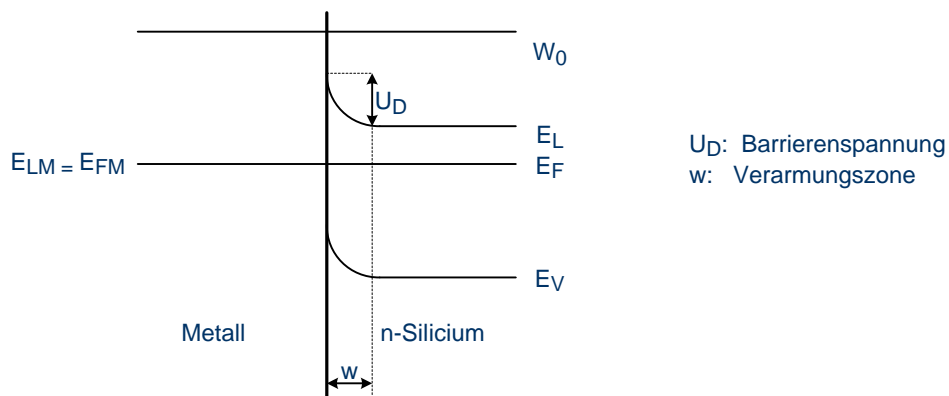


Abb. 5.17: Bändermodell nach dem Kontakt von Metall und n-Silicium

Die Weite w der Verarmungszone hängt von der Stärke der Dotierung ab. Die abgewanderten Elektronen erzeugen im Metall eine negative Raumladung welche auf den Oberflächenbereich begrenzt ist.

Dieser Metall-Halbleiter-Kontakt weist eine nichtlineare Strom-Spannungscharakteristik auf, eine so genannte Schottky-Diode. Diese Barriere können die Elektronen durch Wärmeenergie von außen überwinden oder durch ein anliegendes elektrisches Feld „untertunneln“ (nach der Quantentheorie kann ein Teilchen einen Bereich, in dem es aus energetischen Gründen nicht sein kann, überwinden in dem es sich, stark vereinfacht gesprochen, kurzzeitig Energie ausleiht um die Barriere zu überwinden und die Energie dann wieder zurückgibt: der Tunneleffekt). Auch bei Aluminium kann dieser Effekt beobachtet werden. Da Aluminium an der Oberfläche immer oxidiert, hätten

zwei aneinander liegende Aluminiumflächen eine isolierende Wirkung. Es ist jedoch ein Stromfluss zu verzeichnen, der auf dem Tunneleffekt beruht.

Je nach Anwendung will man diesen Dioden-Effekt herstellen oder aber verhindern. Um einen ohmschen Kontakt, also einen Kontakt ohne diese Potentialbarriere zu erzeugen, kann die Kontaktfläche stark dotiert werden (n^+ -Dotierung), so dass die Verarmungszone sehr dünn wird und der Metall-Halbleiter-Kontakt in Folge des Tunnel-effekts ein lineares Strom-Spannungsverhältnis aufweist.

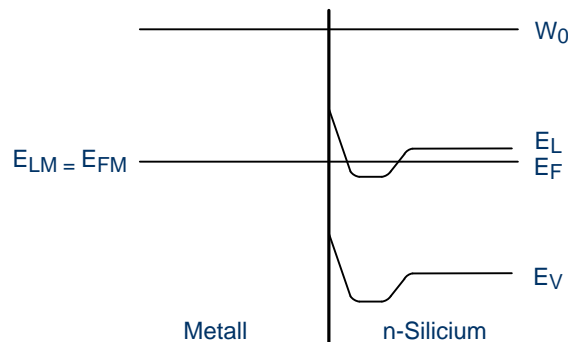


Abb. 5.18: Bändermodell nach n^+ -Dotierung

Da Aluminium im Silicium als Elektronenakzeptor eingebaut wird (es nimmt Elektronen auf) und sich so eine p-Dotierung an der Grenzfläche bildet, entsteht bei p-dotiertem Silicium ein ohmscher Kontakt. Bei einem n-dotierten Gebiet verursacht das Aluminium eine Dotierungsumkehr, so dass hier ein p-n-Übergang entsteht: eine Diode. Um diese zu vermeiden gibt es zwei Möglichkeiten:

- das n-dotierte Gebiet wird so stark dotiert, dass das Aluminium dieses nur abschwächt, nicht aber umkehrt
- eine Zwischenschicht aus Titan, Chrom oder Palladium verhindert die Umdotierung der n-dotierten Gebiete

Zur Verbesserung der Kontakte können auch Metallsilicide (Metalle in Verbindung mit Silicium) an der Kontaktfläche aufgebracht werden.

Bei Metall-p-Halbleiterkontakten ergibt sich in Folge des Ladungsträger austauschs zwischen Metall und Halbleiter eine Bandverbiegung nach unten. Die oben genannten Probleme treten hier nicht auf, da die Elektronen vom energetisch höheren Leitungsband von alleine in das Energieniveau des Metalls gehen.

Im Gegensatz zur Diode beim p-n-Übergang, bei der die Schaltgeschwindigkeit auf

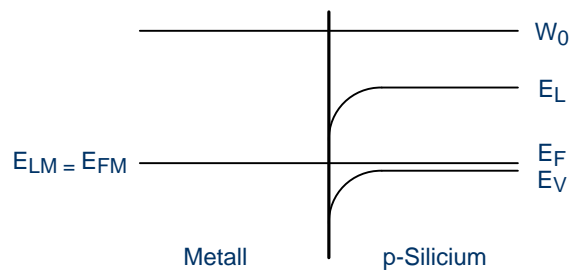


Abb. 5.19: Bändermodell nach dem Kontakt von Metall und p-Silicium

der Diffusion von Elektronen beruht, haben Schottky-Dioden sehr kurze Schaltzeiten. Sie eignen sich daher als Schutzdioden um Spannungsspitzen abzufangen.

5.4.2 Bändermodell eines p-n-Übergangs

Aus der Tatsache, dass das Fermi-niveau konstant sein muss, ergibt sich auch beim p-n-Übergang eine Bänderverbiegung. Diese veranschaulicht die Raumladungszone, welche sich in Folge der abgewanderten Majoritätsladungsträger und der verbleibenden festen Dotieratome einstellt; also die Potentialschwelle, welche im Gleichgewichtszustand (ohne äußere Spannung) eine weitere Diffusion von Elektronen und Löchern in den jeweils anderen Kristall verhindert. Bei Silicium beträgt die Diffusionsspannung zum Überwinden des Potentialgefälles ca. 0,7 V.

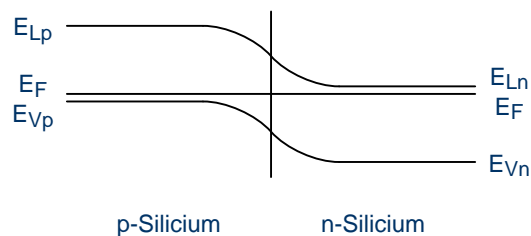


Abb. 5.20: Bändermodell am Übergang p- und n-dotierter Siliciumkristalle

5.5 Mehrlagenverdrahtung

5.5.1 Mehrlagenverdrahtung

Die Verdrahtung kann in einer integrierten Schaltung über 80 % der Chipfläche einnehmen, darum wurden Techniken entwickelt, mit denen man die Verdrahtung in mehrere Ebenen übereinander legt. So lässt sich die Summe der Leiterbahnen bei nur einer zusätzlichen Ebene um bis zu 30 % verringern. Dennoch kann die Gesamtlänge aller Leiterbahnen in modernen Mikrochips bis weit über 10 km(!) betragen.

Zwischen den Verdrahtungsebenen sind Isolationsschichten aufgebracht, durch Kontaktöffnungen (VIA, vertical interconnect access) werden die einzelnen Ebenen miteinander verbunden. Dabei sind heute sieben und mehr Verdrahtungsebenen gebräuchlich. Steile Kanten und Stufen müssen entschärft werden, da die Konformität der aufgetragenen Metallisierungen gering ist und somit Engstellen entstehen, die wiederum durch sehr hohe Stromdichten belastet werden. Folge: die Leiterbahnen altern frühzeitig oder reißen ab. Um die Kanten bzw. Stufen zu entfernen gibt es mehrere Möglichkeiten zur Planarisierung, die im folgenden erläutert werden.

5.5.2 BPSG-Reflow

Bei der Reflowtechnik werden Schichten aus dotierten Gläsern auf dem Wafer aufgebracht. Weit verbreitet sind Phosphorsilicatglas (PSG) und Borphosphorsilicatglas (BPSG). In einem Hochtemperaturschritt verfließen die Gläser und bilden eine ebene Oberfläche: bei PSG und BPSG geschieht dies bei ca. 900 °C. Zur Planarisierung auf einer Verdrahtungsebene ist diese Technik jedoch nicht geeignet, da das Aluminium unter den hohen Temperaturen aufschmelzen würde.

5.5.3 Reflowrückätzen

Auf dem Wafer wird eine Siliciumdioxidschicht aufgebracht, die mindestens so dick ist wie die höchste Stufe auf der Scheibe. Auf der Oxidschicht wird eine Lack- oder Polyimidschicht aufgeschleudert und zur Stabilisierung thermisch behandelt (siehe Fototechnik); durch die Temperatur verfließt die Schicht.

Im Trockenätzverfahren werden der Lack bzw. das Polyimid und das Siliciumdioxid

mit gleichen Ätzraten abgetragen, so dass eine eingeebnete Oxidoberfläche zurückbleibt.

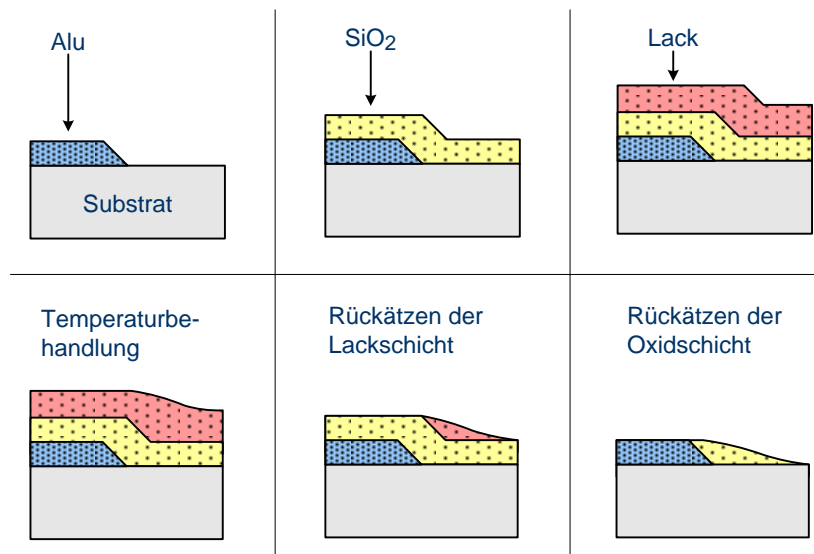


Abb. 5.21: Prozessschritte beim Reflowrückätzen

Neben der Technik mit Lack oder Polyimid kann auch so genanntes Spin On Glas (SOG) auf dem Wafer aufgebracht werden. Ebenfalls im Schleuderverfahren entsteht so direkt eine planarisierte Schicht, die durch eine thermische Behandlung stabilisiert wird. Die vorhergehende Siliciumdioxidschicht ist hierbei nicht erforderlich. Diese Techniken bieten jedoch keine Homogenität über den gesamten Wafer, sondern gleichen nur lokal Stufen aus.

5.5.4 Chemisch-Mechanisches Polieren

Beim chemisch mechanischen Polieren (auch chemisch mechanisches Planarisieren, kurz CMP) wird im Gegensatz zu den Reflowtechniken eine Homogenität über die gesamte Scheibenoberfläche erzielt. Dies ist vor allem im Hinblick auf lithografische Prozesse wichtig, die zur korrekten Belichtung eine möglichst planare Oberfläche benötigen. Ebenso ist eine Waferoberfläche ohne Topografie für alle nachfolgenden Schichten von Vorteil.

Der Wafer wird dazu in einem Chuck mit Vakuumansaugung (Head) mit der aktiven Seite nach unten gehalten und auf eine Polierfläche (Pad, meist aus Polyurethan) auf dem Poliertisch gepresst. Der Head und der Poliertisch drehen sich, während der Head

gleichzeitig horizontale Bewegungen ausführen kann. Als Poliermittel zwischen Tisch und Wafer dient eine Lösung (Slurry) aus Schleifmitteln und chemischen Substanzen, die unter Druck die Oberfläche verändern und so den Polierprozess unterstützen. Zur besseren Verteilung der Slurry und um das Poliertuch zur konditionieren, kann das Pad mit einer mit Diamanten besetzten Stahlscheibe (Dresser/Conditioner) aufgeraut werden. Dies geschieht während (in-situ) oder vor/nach dem Polierschritt (ex-situ).

Der CMP-Prozess erfolgt gewöhnlich in zwei oder drei Stufen auf Pads mit unterschiedlichen Oberflächen und verschiedenen Slurrys. Die Wafer werden dazu nach jedem Polierschritt auf das nächste Pad transferiert. Im Anschluss erfolgt eine Reinigung, um Partikel und Slurryreste zu entfernen.

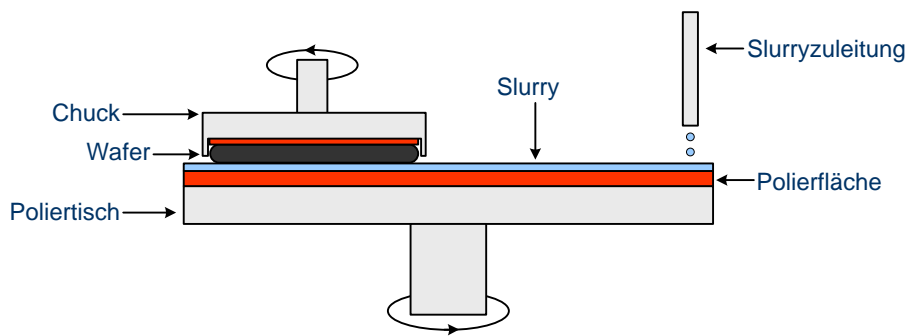


Abb. 5.22: Schematische Darstellung einer CMP-Anlage

Typischerweise wird der CMP-Prozess nach der Abscheidung des TEOS zur Shallow-Trench-Isolation eingesetzt um das Oxid soweit abzutragen, dass nur die Isolation zwischen den aktiven Gebieten der Transistoren bestehen bleibt. Ebenso wird das Zwischenoxid zwischen Transistorebene und der ersten Metallebene (First contact) mittels CMP auf die erforderliche Dicke zurückpoliert. In diesem Oxid werden anschließend die Kontakte zu den Source- und Draingebieten mittels Wolfram hergestellt. Auch hier dient das chemisch mechanische Polieren dazu, das auf der Oberfläche befindliche Metall zu entfernen. Wie im Kapitel Damscene-Verfahren beschrieben wurde, werden auch die Verdrahtungsebenen aus Kupfer im CMP-Prozess eingeebnet.

Im Folgenden ist der CMP-Prozess mit zwei Polierschritten im STI-Bereich dargestellt. Nach dem ersten Polierschritt wird das Oxid auf dem aktiven Gebiet und über den Gräben planarisiert. Im zweiten Polierschritt wird das restliche Oxid dann in einem selektiven Prozess bis auf die Passivierungsschicht abgetragen. Hierbei ist wichtig, dass das Oxid auf den Bereichen, auf denen später die Transistoren hergestellt werden, vollständig entfernt wird, da sonst das Nitrid, welches das darunterliegende Silicium

während des Polierens schützt, nicht nasschemisch entfernt werden kann.

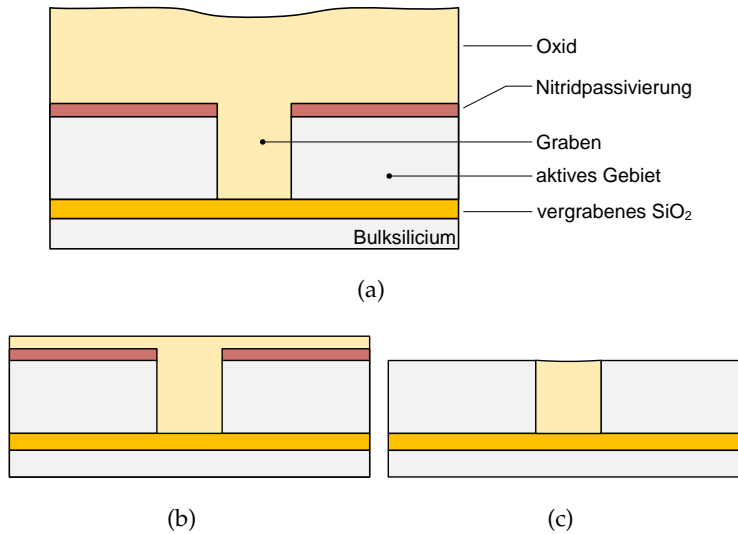


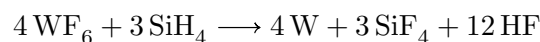
Abb. 5.23: CMP-Prozess zur Grabenisolation: (a) Vor dem Polieren, (b) nach dem ersten Polierschritt, (c) nach dem finalen Polierschritt und Nitridentfernen

Auch wenn dieses Verfahren recht grob anmutet, lässt sich hierbei doch eine auf wenige Nanometer planare Oberfläche herstellen.

5.5.5 Kontaktierung der Metallisierungsebene

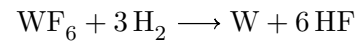
Um die Metallebenen zu verbinden, werden in die Isolationsschichten Kontaktlöcher mit sehr hoher Anisotropie geätzt, so werden Kanten an den Kontaktlöchern vermieden. Die Kontaktlöcher müssen so aufgefüllt werden, dass einerseits eine optimale Kontaktierung gewährleistet wird und zugleich die Oberfläche planar bleibt.

Zur Auffüllung der Kontaktlöcher hat sich Wolfram als geeignet erwiesen. Unter Zugabe von Silan scheidet sich in einem CVD-Prozess aus Wolframhexafluorid eine dünne Schicht Wolfram als Nukleationskeim ab, Siliciumtetrafluorid und Fluorwasserstoff als Nebenprodukte werden abgesaugt:



Unter Zugabe von Wasserstoff zu Wolframhexafluorid werden die Kontaktlöcher dann

aufgefüllt:



Darüber kann die nächste Metallebene aufgebracht, strukturiert und planarisiert werden. Bei Kupfer als Metallisierung wird Wolfram nur für den ersten Kontakt zum Siliciumsubstrat benötigt. Die Verbindung der einzelnen Kupferebenen geschieht mit dem Metall selbst.

6 Lithografie

6.1 Belichten und Belacken

6.1.1 Übersicht

In der Halbleiterfertigung werden Strukturen auf Siliciumscheiben mittels lithografischer Verfahren hergestellt. Dabei wird zuerst ein strahlungsempfindlicher Film, meist eine Fotolackschicht, auf dem Wafer aufgebracht, strukturiert und mit Ätzverfahren in die darunter liegende Schicht übertragen.

Die Fototechnik beinhaltet dabei folgende Prozessschritte:

- Aufbringen eines Haftvermittlers und Entfernen von Wasser auf dem Wafer
- Belacken der Wafer
- Stabilisieren der Lackschicht
- Belichten
- Entwickeln des Lackes
- Aushärten des Lackes
- Kontrolle

Bei einigen Prozessen wie der Ionenimplantation, dient der Fotolack als Schutzschicht um bestimmte Bereiche auf dem Wafer von der Implantation auszuschließen. Eine Übertragung der Lackmaske durch einen Ätzprozess findet hier nicht statt.

6.1.2 Aufbringen eines Haftvermittlers

Zu Beginn werden die Wafer gereinigt und ausgeheizt (Pre-Bake) um anhaftende Partikel zu beseitigen und angelagertes Wasser zu entfernen. Die Oberfläche der Wafer ist

Wasser anziehend (hydrophil) und muss vor dem Aufbringen der Lackschicht hydrophob, also Wasser abstoßend und damit Lack anziehend gemacht werden. Dazu wird auf den Wafern ein Haftvermittler, meist Hexamethyldisilazan (HMDS), aufgebracht. Die Wafer werden dabei dem Dampf dieser Flüssigkeit ausgesetzt, so dass sich die Scheibenoberflächen damit benetzen.

Durch die Feuchtigkeit in der Umgebungsluft befinden sich auch nach dem Ausheizen immer Wasserstoff H oder Hydroxidionen OH^- an der Waferoberfläche. Das HMDS spaltet sich in Trimethylsiliciumgruppen $\text{Si}-3\text{CH}_3$ auf und entfernt den Wasserstoff unter Bildung von Ammoniak NH_3 .

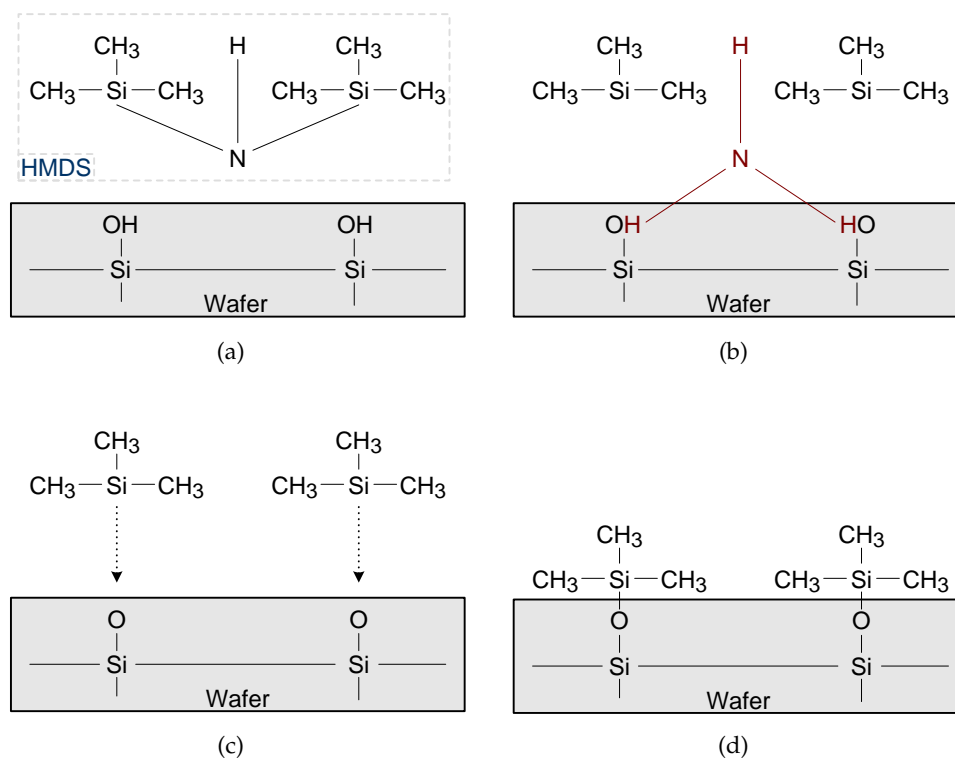


Abb. 6.1: Oberflächenmodifikation mit HMDS

6.1.3 Belacken

Die Belackung der Wafer erfolgt durch eine Schleuderbeschichtung auf einem drehbaren Teller mit Vakuumsaugung (Chuck). Bei niedriger Drehzahl wird Lack in der Mitte der Scheibe aufgespritzt und dann bei 2000-6000 Umdrehungen pro Minute

durch die Zentrifugalkraft zu einer homogenen Lackschicht auseinander gezogen. Deren Dicke beträgt je nach anschließendem Prozess bis zu $2\text{ }\mu\text{m}$. Die Dicke hängt dabei von der Drehzahl und der Zähigkeit (Viskosität) des Lackes ab.

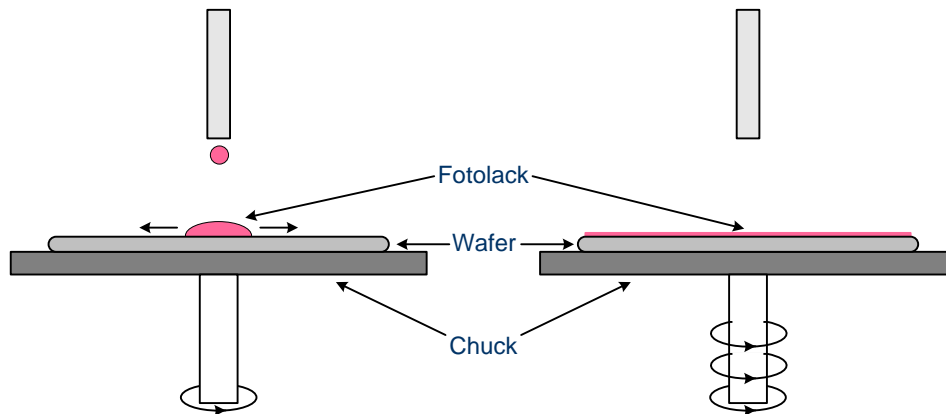


Abb. 6.2: Schleuderbelackung

Damit der Lack auf der Scheibe gleichmäßig verfließen kann enthält er Wasser und Lösungsmittel, die ihn weich machen. Zur Stabilisierung der Lackschicht wird der Wafer danach bei ca. $100\text{ }^{\circ}\text{C}$ ausgeheizt (Post-Bake bzw. Soft-Bake). Wasser und Lösungsmittel werden teilweise verdampft, eine Restfeuchtigkeit muss für die Belichtung erhalten bleiben.

6.1.4 Belichtung

In einer Belichtungsanlage befindet sich eine Glasmaske die teilweise mit Chrom beschichtet ist, dadurch werden partiell Bereiche des belackten Wafers belichtet und andere nicht.

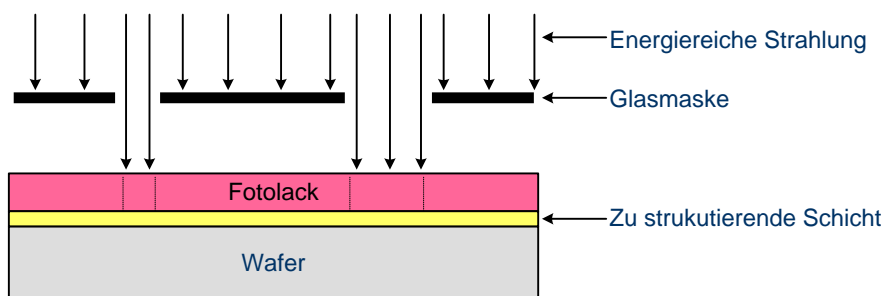


Abb. 6.3: Prinzip der optischen Fotolithografie

Je nach Art des Lackes werden belichtete Teile löslich oder unlöslich. Mit Hilfe einer Entwicklerlösung werden die löslichen Bereiche entfernt, so dass eine strukturierte Lackschicht erhalten bleibt. Bei Positivlack spaltet sich ein Stickstoffmolekül (N_2) durch das energiereiche UV-Licht ab. Zurück bleibt ein so genanntes Keto-Karben, das sich aus energetischen Gründen zu Keten (Strukturformel $CH_2=C=O$) umwandelt. Unter Aufnahme von Feuchtigkeit aus der Umgebungsluft bildet sich aus dem Keten eine Carbonsäure.

Die Belichtungszeit ist sehr wichtig damit die Strukturen die exakte Größe erhalten. Je länger die Scheibe den Strahlen der Belichtungsanlage ausgesetzt wird, desto größer werden dabei die belichteten Bereiche. Eine exakte Bestimmung der korrekten Belichtungsdauer zum Erreichen der vorgegebenen Strukturbreiten mit einem oder mehreren Testwafern (Vorläufern) ist notwendig, da sich der Lack je nach Umgebungstemperatur auch unterschiedlich verhalten kann.

Bei einer Überbelichtung werden Lackstege und damit die darunter liegenden Strukturen zu klein, Kontaktlöcher werden zu groß. Bei einer zu kurzen Belichtungszeit sind die Kontaktlöcher nicht geöffnet, Leiterbahnen sind zu breit und stehen unter Umständen miteinander in Kontakt. Zudem führt eine schlechte Fokussierung zu unbelichteten Bereichen, so dass Kontaktlöcher bei der Entwicklung später nicht freigelegt werden und zwischen Leiterbahnen Verbindungen bestehen bleiben, die zu Kurzschlüssen führen.

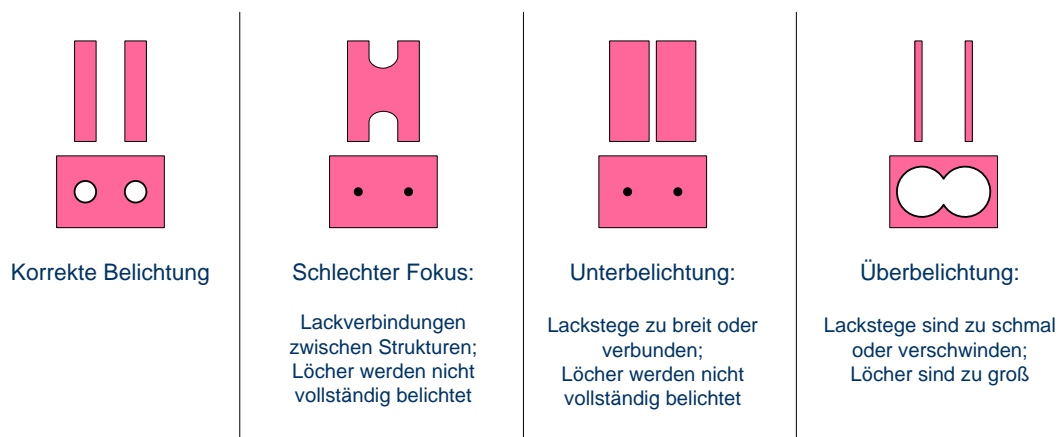


Abb. 6.4: Fehler bei der Belichtung durch falschen Fokus, Über- oder Unterbelichtung

Je nach nachfolgendem Prozess muss das Lackmaß, also die Breite der Lackstege oder der Durchmesser der Löcher angepasst werden. Bei isotropen Ätzungen (die Ätzung

geschieht sowohl in vertikaler als auch horizontaler Richtung) wird die Maskierung nicht 1:1 in die zu strukturierende Schicht übertragen.

6.1.5 Eingesetzte Belichtungsverfahren

Zur Belichtung werden je nach Anforderung energiereiche Strahlen wie UV-Licht, Elektronenstrahlen, Röntgenstrahlung und Ionenstrahlen eingesetzt. Dabei gilt, je kürzer die Wellenlänge der Strahlung, desto kleiner sind die erreichbaren Strukturbreiten.

Bei 120 nm Strukturen wird beispielsweise ein Argon-Fluorid-Laser eingesetzt (ArF-Laser), für kleinere Strukturen Stickstoff-Laser (Wellenlänge 155 nm) oder Extreme UV-Strahlung (EUV, Wellenlänge 13 nm). Röntgenstrahlen haben eine Wellenlänge von 0,2-0,4 nm, Elektronenschreiber ca. 0,02 nm und Ionenstrahler, im Falle von Wasserstoffionen 0,0001 nm.

Dass man mit der Wellenlänge von x nm kleinere Strukturen als x nm herstellen kann liegt daran, dass man sich spezieller Belichtungsmasken bedient, mit denen sich eine Phasenverschiebung des Lichts ausnutzen lässt. Zudem kann das Auflösungsvermögen mit Flüssigkeitsfilmen (u.a. Reinstwasser) im Linsensystem erhöht werden (Immersion). Durch diese Techniken können auch heutige Strukturen von nur mehr 45 nm nach wie vor mit einer Belichtungswellenlänge von 193 nm hergestellt werden.

Während UV-Strahlung (erzeugt durch Quecksilberdampflampen) und Gaslaser heutzutage bei der Belichtung von Wafern Anwendung finden, werden Röntgen- und Ionenstrahlen meist für Forschungszwecke eingesetzt. Elektronenstrahlschreiber werden zur Herstellung der Fotomasken verwendet.

6.2 Belichtungsverfahren

6.2.1 Übersicht

Die Fototechnik kann, je nach Art der Bestrahlung in mehrere Verfahren unterteilt werden: optische Lithografie (Fotolithografie), Elektronenstrahllithografie, Röntgenstrahllithografie und zukünftig Ionenstrahllithografie.

Bei der optischen Lithografie werden strukturierte Masken oder Reticle verwendet. Die Belichtung mit UV-Licht oder Gaslasern erfolgt entweder im Maßstab 1:1 oder

reduzierend, beispielsweise 4:1 oder 10:1.

6.2.2 Kontaktbelichtung

Das Kontaktbelichtungsverfahren ist das älteste angewandte Verfahren. Dabei wird die Maske direkt auf die Lackschicht gepresst, die Strukturen im Maßstab 1:1 übertragen. Auflösungsbegrenzende Streu- bzw. Beugungseffekte des Lichts treten nur an den Strukturkanten auf. Die erzielten Strukturweiten sind bei diesem Verfahren jedoch gering. Da alle Chips auf einmal belichtet werden ist der Scheibendurchsatz bei dieser Technik sehr hoch, der Aufbau der Belichtungsanlagen ist relativ einfach.

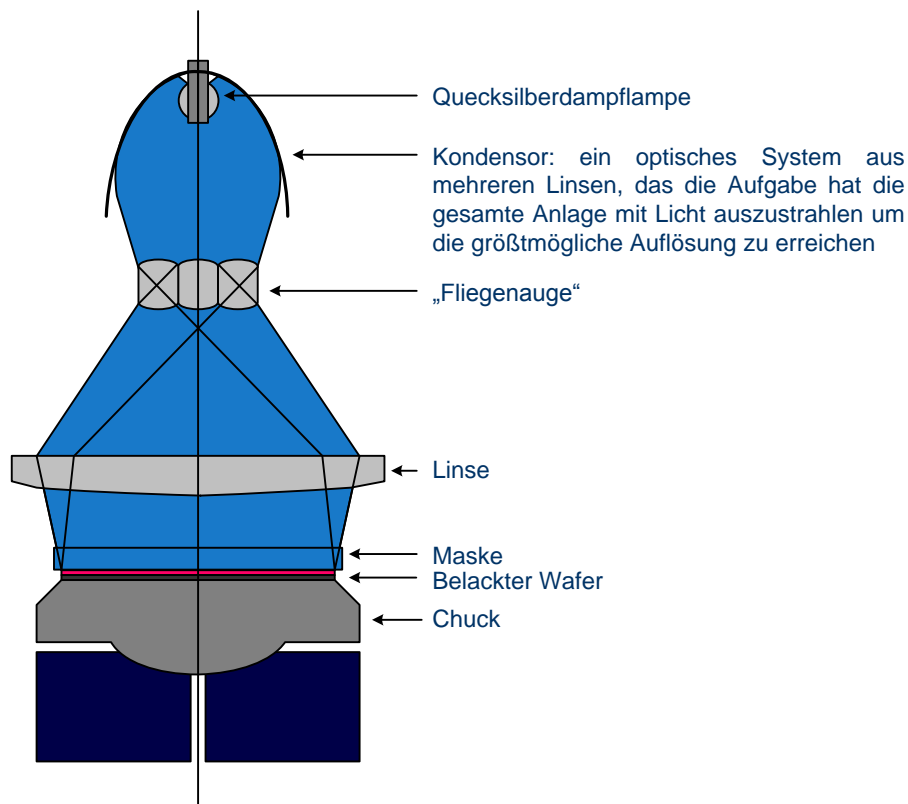


Abb. 6.5: Kontaktbelichtung

Die Nachteile liegen jedoch auf der Hand: durch die angepresste Maske auf den Lack verschmutzt diese schnell, oder kann, wie auch der Lack, dabei verkratzt werden. Befinden sich Partikel zwischen Scheibe und Maske wird die Abbildung durch den Abstand von Maske und Wafer verschlechtert.

6.2.3 Abstandsbelichtung

Bei der Abstands- oder Proximitybelichtung wird der Kontakt von Maske und Scheibe durch einen Abstandshalter (ca. 20 μm) vermieden. Dabei wird jedoch nur ein Schattenbild der Maske auf dem Wafer abgebildet, welches eine deutlich schlechtere Auflösung der Strukturen bietet.



Abb. 6.6: Abstandsbelichtung

6.2.4 Reduzierende Projektionsbelichtung

Bei diesem Verfahren ist die Step-and-Repeat-Technik gebräuchlich. Dabei wird ein einzelner Chip - bei geringer Größe auch mehrere - über ein Reticle auf dem Wafer abgebildet. So wird die gesamte Fläche der Scheibe Chip für Chip belichtet.

Der Vorteil bei diesem Verfahren ist, dass die auf dem Reticle abgebildeten Strukturen um den Faktor 4 oder 10 vergrößert vorliegen. Bei der verkleinerten Abbildung auf den Wafer werden neben den Strukturen auch alle Fehler, wie Partikel, verkleinert dargestellt oder fallen sogar unter die Auflösungsgrenze; die Auflösung selbst wird bei diesem Verfahren verbessert. Da für eine Ebene nicht die komplette Maske genutzt wird, können mehrere Ebenen auf einer Glasplatte aufgebracht werden, so dass die Maskenkosten gesenkt werden.

Zusätzlich kann ein Pellicle, eine dünne Folie mit Alurahmen, auf die Maske geklebt werden, wodurch Partikel von der Maske ferngehalten werden. Diese befinden sich nun außerhalb des Schärfebereichs und werden nicht abgebildet.

Daneben gibt es noch die Spiegelprojektionsbelichtung (1:1), bei der der Wafer über ein komplexes System aus Spiegeln belichtet wird. Durch die Verwendung von Spiegeln entstehen keine Farbfehler, wie sie bei Linsensystem auftreten, dazu können Ausdehnungen der Scheibe durch Temperatureinfluss in Prozessen ausgeglichen werden. Jedoch werden über Spiegelsysteme Bilder verzerrt/gekrümmt dargestellt. Durch die 1:1-Abbildung ist die Auflösung außerdem stark begrenzt.

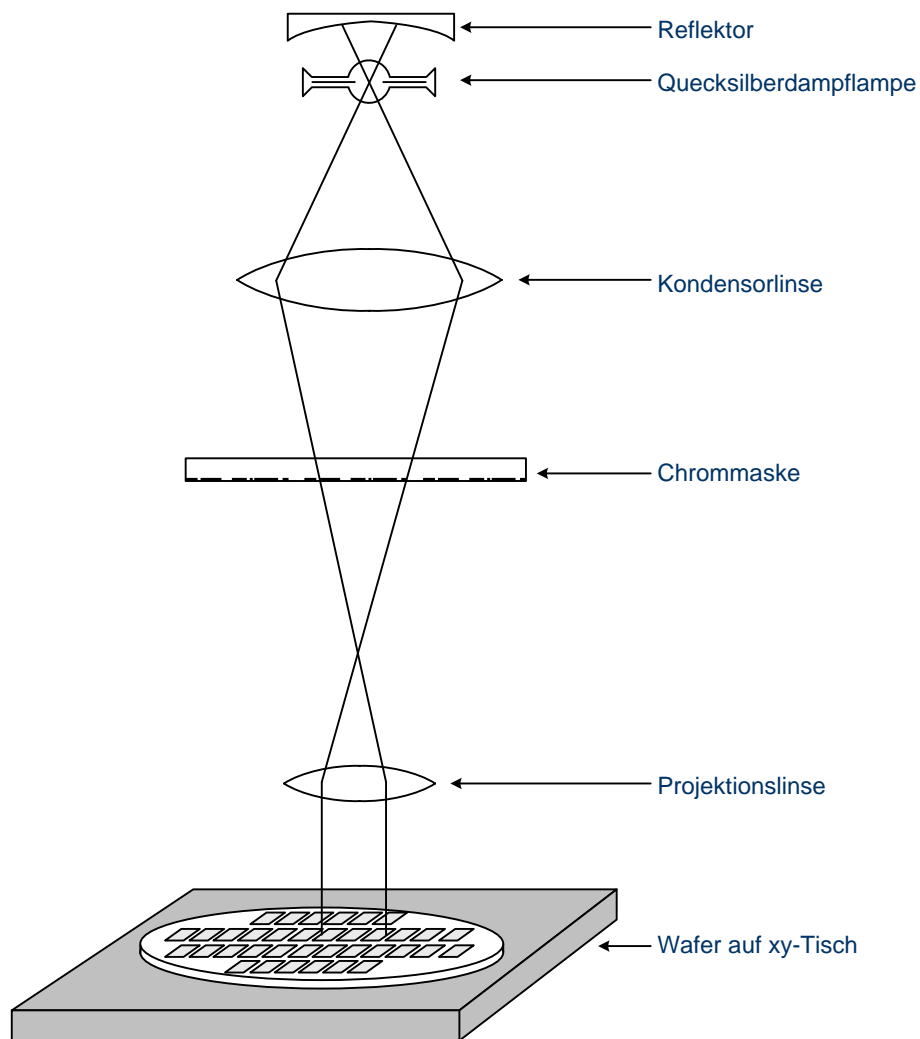


Abb. 6.7: Step-and-Repeat-Belichtung

6.2.5 Elektronenstrahlithografie

Wie bei der Maskenherstellung wird ein fokussierter Elektronenstrahl über den belackten Wafer gescannt. Dabei kann das Scannen zeilenweise im Raster-Scan-Verfahren oder im Vektorscanverfahren durchgeführt werden. Jedoch muss auch hier jede Struktur einzeln geschrieben werden, was sehr zeitintensiv ist. Der Vorteil ist, dass keine Masken benötigt werden, was Kosten spart. Die gesamte Apparatur befindet sich dabei im Hochvakuum.

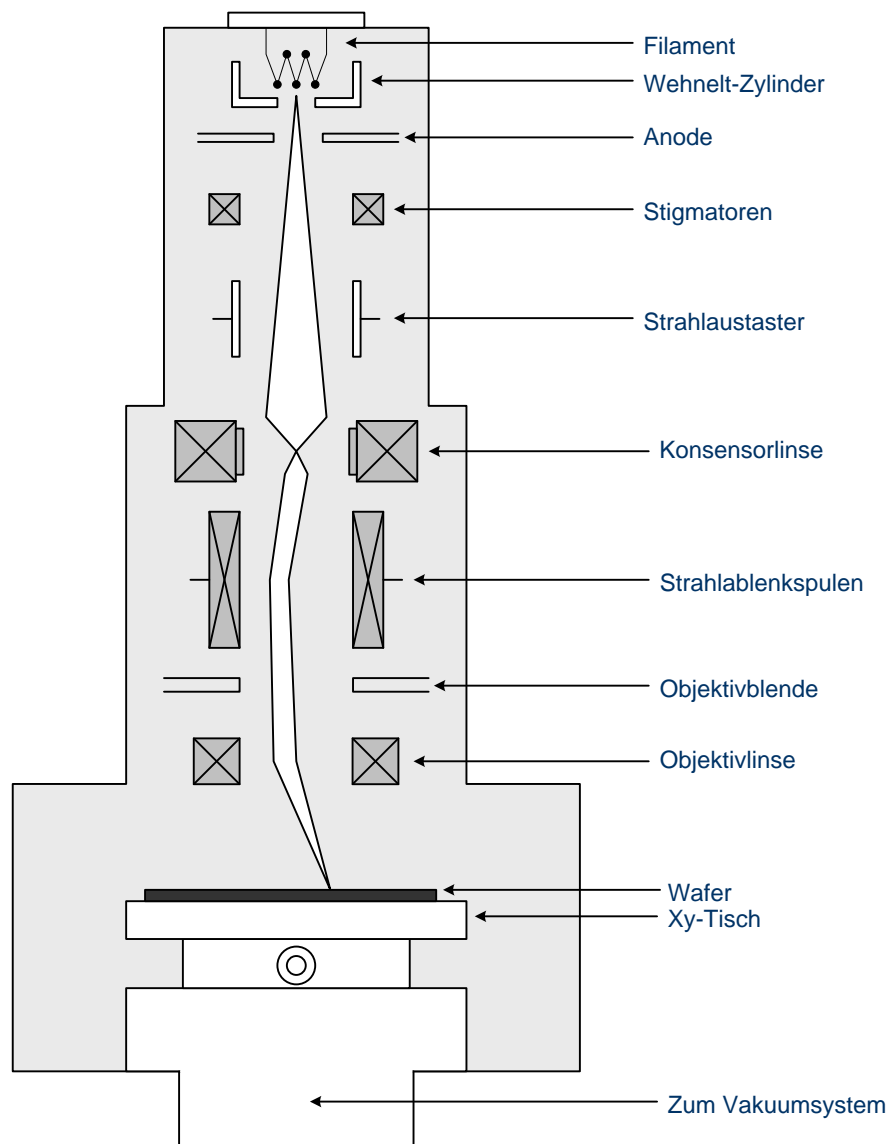


Abb. 6.8: Elektronenstrahlolithografie

6.2.6 Röntgenstrahlolithografie

Die Auflösungsgrenze der Röntgenstrahlolithografie liegt bei etwa 40 nm. Die Abbildung erfolgt im 1:1 Step-and-Repeat-Verfahren durch Schattenwurf, und wird unter Atmosphärendruck in Luft oder bei leichtem Unterdruck in Heliumatmosphäre (ca. 10.000 Pa) durchgeführt. Die Röntgenquelle kann dabei eine Plasmaquelle oder Synchrotronstrahlung sein.

Anstelle der chrombeschichteten Glasmasken werden dünne, mechanisch stabile Folien aus Beryllium, teils auch aus Silicium verwendet. Um die Röntgenstrahlung zu absorbieren werden die Folien mit schweren Elementen wie Gold beschichtet. Die Anlagen sowie die Masken sind sehr teuer.

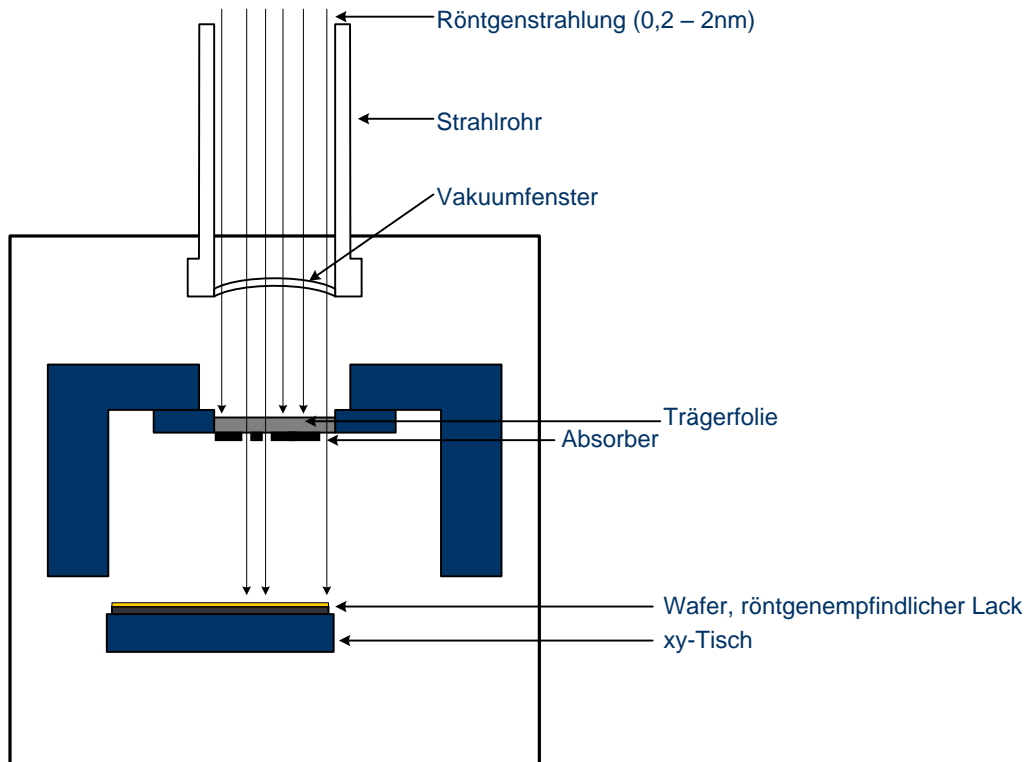


Abb. 6.9: Röntgenstrahlolithografie

6.2.7 Weitere Verfahren

Eine weitere Möglichkeit der Lithografie ist die Bestrahlung der Wafer mit Ionen. Mit den Ionen kann der Wafer sowohl über eine Maske strukturiert, als auch direkt wie bei der Elektronenstrahlmethode beschrieben werden. Im Falle von Wasserstoffionen beträgt die Wellenlänge 0,0001 nm. Mit anderen Elementen ist auch eine direkte Dotierung ohne Maskierung denkbar.

6.3 Der Fotolack

6.3.1 Lacktechnik

Es gibt Positiv- und Negativlacke, die bei unterschiedlichen Anwendungen eingesetzt werden. Während beim Positivlack die belichteten Stellen löslich werden, werden bestrahlte Bereiche beim Negativlack für die Entwicklerlösung unlöslich.

Charakteristik der Positivlacke:

- + sehr gute Auflösung
- + stabil in der Entwicklerlösung
- + in Laugenlösungen entwickelbar
- nur bedingt resistent gegen Ätzung und Implantationsprozesse
- schlechte Haftung auf dem Wafer

Charakteristik der Negativlacke:

- + sehr empfindlich (exaktere Strukturierung bei der Belichtung möglich als Positivlack)
- + gute Haftung
- + resistent gegen Ätzverfahren und Implantation
- + billiger als Positivlack
- geringe Auflösung
- Xylol-Entwickler (giftig)

Zur Strukturierung der Wafer in der Fertigung werden fast ausschließlich Positivlacke verwendet. Negativlacke finden meist als Passivierungsschicht (Fotoimidschicht) Anwendung, die mittels UV-Licht ausgehärtet werden können. Sofern im Text keine Angaben zum Lack gemacht werden, handelt es sich um Positivlack; bei Negativlack wird dies ausdrücklich erwähnt.

6.3.2 Chemische Zusammensetzung

Fotolacke (auch als Fotoresist bezeichnet; resist = widerstehen) setzen sich aus einem Bindemittel, einem Sensibilisator und einem Lösemittel (Verdünner) zusammen.

- **Bindemittel (Anteil 20 %):** Als Bindemittel wird häufig Novolac eingesetzt. Dabei handelt es sich um ein Phenolharz (Kunstharz, Kunststoff) und bestimmt in erster Linie die thermischen Eigenschaften des Lackes.
- **Sensibilisator (Anteil 10 %):** Der Sensibilisator bestimmt die Lichtempfindlichkeit des Lackes. Sensibilisatoren setzen sich aus Molekülen zusammen, die bei einer Belichtung mit energiereicher Strahlung die Löslichkeit des Lackes verändern (im Positivlack entsteht aus dem Sensibilisator durch die Belichtung Carbonsäure. Mehr dazu im Kapitel Entwickeln). Damit der Lack nicht durch das Licht in den Fertigungshallen belichtet wird, finden die Fototechnikprozesse unter Gelblicht statt, gegen das der Lack unempfindlich ist.
- **Lösemittel (Anteil 70 %):** Die Lösemittel bestimmen die Viskosität des Lackes. Durch das Verdampfen der Lösemittel auf einer Heizplatte wird der Lack stabilisiert und resistent für nachfolgende Prozesse.

Ein vom Hersteller gelieferter Lack hat eine definierte Oberflächenspannung und Dichte, einen bestimmten Feststoffgehalt und eine bestimmte Viskosität. Somit ist bei der Belackung in der Chipfertigung die erzielte Fotolackdicke ausschließlich von der Drehzahl der Belackungsanlage abhängig.

6.4 Entwickeln und Kontrolle

6.4.1 Entwickeln

Die belichteten Scheiben werden in Tauch- oder Sprühentwicklungen prozessiert. Beim Tauchätzschritt wird eine komplette Horde in einer Laugenlösung entwickelt und anschließend gespült, bei der Sprühentwicklung wird jeder Wafer einzeln bearbeitet. Wie beim Belacken befindet sich der Wafer auf einem Chuck und wird bei langsamer Umdrehung stetig mit Entwicklerlösung besprüht. Nach dem vollständigen Entwickeln des Lackes wird der Wafer mit aufgespritztem Wasser gespült um den Entwicklungsprozess zu stoppen. Einige Vorteile der Sprühentwicklung gegenüber dem Tauchver-

fahren sind folgende:

- kleinste Strukturen können freigelegt werden
- Die Entwicklerlösung wird stetig erneuert: Verunreinigungen werden verhindert
- Die Menge der eingesetzten Entwicklerlösung ist wesentlich geringer

Durch das Entwickeln lösen sich, entsprechend der Lackart, bestimmte Bereiche des Lackes, so dass am Ende eine strukturierte Scheibe bestehen bleibt. Durch die Belichtung wird im Lack eine Reaktion ausgelöst, bei der der Sensibilisator in Säure umgewandelt wird. Diese Carbonsäure wird beim Entwickeln nach folgender Gleichung mit Natronlauge (NaOH) zu wasserlöslichem Salz umgewandelt:

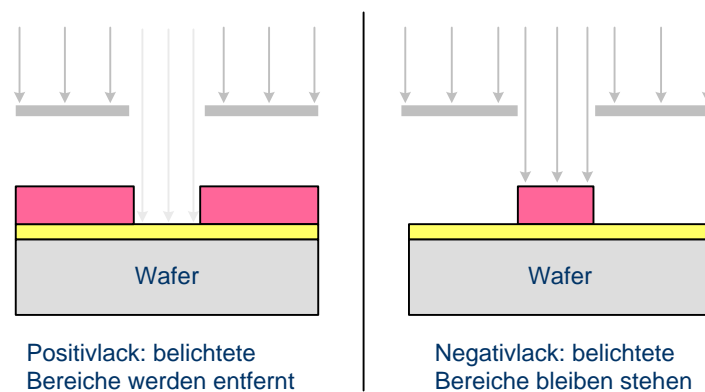
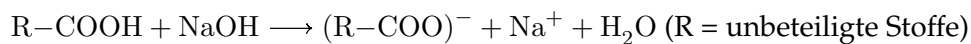


Abb. 6.10: Darstellung der belichteten Lackarten nach dem Entwickeln

Da bei Kali- oder Natronlösungen Rückstände des Entwicklers auf dem Wafer zurückbleiben, werden auch Metallionen freie Entwickler wie TMAH (Tetramethylammoniumhydroxid) eingesetzt. Durch einen erneuten Aushärteschritt (Hard-Bake) wird der Lack für nachfolgende Prozesse, wie Ätzen oder für eine Ionenimplantation, beständig gemacht.

6.4.2 Lackkontrolle

Nun wird die Lackstruktur kontrolliert. Unter schrägem Lichteinfall lassen sich im Mikroskop die Gleichmäßigkeit der Lackschicht, sowie schlechte Fokussierungen und

Lackanhäufungen erkennen. Sind die Lackstege zu dick oder dünn muss der Lack entfernt und der Prozess wiederholt werden. Ebenso muss die Lackstruktur zu der sich darunter befindlichen Ebene exakt justiert sein, andernfalls ist ebenso eine Wiederholung der Belackungs- und Belichtungsprozesse notwendig. Dazu gibt es unterschiedliche Justiermarken für die Justiergenauigkeit und die Linienweitenkontrolle:

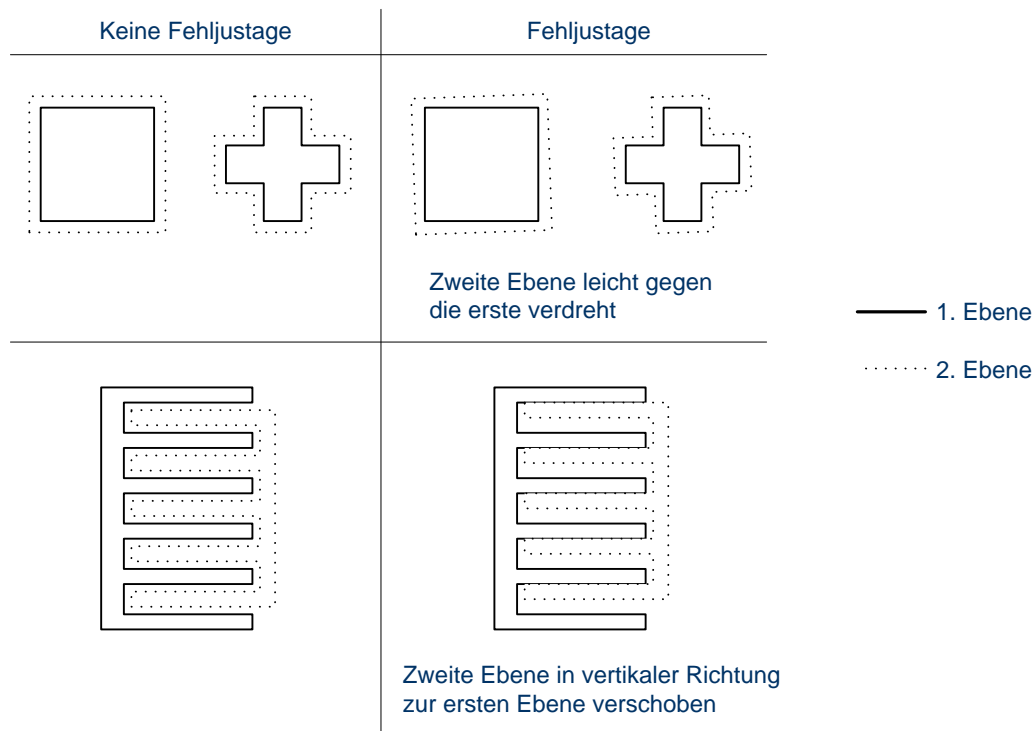


Abb. 6.11: Darstellung von Justiermarken

Die Breite der Lackstege wird mit einem Mikroskop kontrolliert: Lichtstrahlen treffen senkrecht auf den Wafer, und werden an Kanten nicht zum Objektiv zurückgestreut. So sind schwarze Linien als Strukturbegrenzungen sichtbar, mit deren Abständen zueinander man unter Zuhilfenahme der Mikroskopvergrößerung die Breite der Stege berechnen kann.

6.4.3 Lackentfernen

Nachdem die Struktur unter dem Lack in Ätzverfahren abgetragen wurde, oder der Lack beim Implantationsprozess als Maskierungsschicht gedient hat, muss der Lack entfernt werden. Dies geschieht mit starken Ätzlösungen (Remover), in einem Tro-

ckenätzschritt oder mit Lösungsmitteln. Als Lösungsmittel zum Entfernen der Lackschicht eignet sich Aceton, da es den Wafer und andere Schichten nicht angreift. Durch eine Ionenimplantation oder einen Trockenätzprozess kann die Lackschicht jedoch so stark ausgehärtet sein, so dass Lösungsmittel den Lack nicht mehr angreifen und entfernen können.

In diesem Fall kann der Lack mit einem Remover bei ca. 80 °C in einem Tauchverfahren entfernt werden. Hat sich der Lack während der Bearbeitung auf über 200 °C erwärmt kann er auch vom Remover nicht mehr abgetragen werden. Dann muss der Lack mittels Lackveraschung oder Lackverbrennung entfernt werden.

Unter Anwesenheit von Sauerstoff wird dabei durch Hochfrequenzanregung eine Gasentladung gezündet, so dass angeregte Sauerstoffatome entstehen. Diese verbrennen den Lack rückstandsfrei. Die geladenen Teilchen werden jedoch durch das elektrische Feld stark beschleunigt und können so einen leichten Abtrag der Scheibenoberfläche oder eine geringe Schädigung der Scheibe verursachen.

6.5 Maskentechnik

6.5.1 Maskentechnik

Die in der Fototechnik eingesetzten Masken enthalten ein Muster mit dem die jeweilige Schicht auf dem Wafer strukturiert wird. Ausgangsmaterial für die Masken sind Glasplatten, die ganzflächig mit Chrom und Lack beschichtet sind. Über den für Elektronenstrahlen empfindlichen Lack wird die Chromschicht strukturiert, welche dann die lichtundurchlässigen Bereiche auf der Glasmaske darstellt.

Mit einem Elektronenstrahl werden die Masken direkt beschrieben. Die gesamte Apparatur – die Elektronenstrahlquelle, Fokussier- und Ablenkeinheit und die Glasplatte – befindet sich im Hochvakuum (0,01-100 Pa; normaler Luftdruck ca. 100.000 Pa). Der Elektronenstrahl wird computergesteuert über die Maske gelenkt und belichtet den Lack. Bei dieser Methode lassen sich Strukturen bis weit unter 100 nm auflösen.

6.5.2 Schritte bei der Maskenherstellung

Grundsätzlich werden die Strukturen auf den Masken auf die gleiche Weise hergestellt, wie auf den Wafern. Im Gegensatz zum Belichtungsprozess in der Waferfertigung, bei dem die Strukturen der Maske durch Schattenwurf auf dem Lack abgebildet werden, werden die Masken aber mit einem Elektronenstrahl direkt beschrieben.

1. Belichten eines fotoempfindlichen Lacks zur Strukturierung einer Chromschicht auf dem Glassubstrat mittels Laser oder Elektronenstrahl

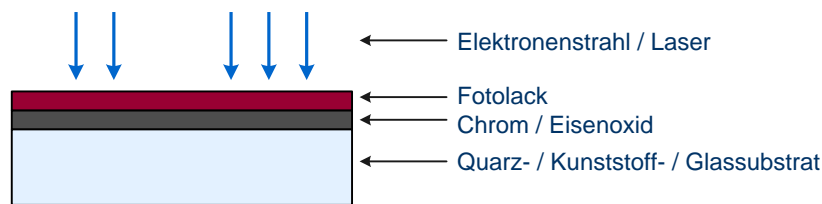


Abb. 6.12: Prinzipieller Schichtaufbau einer Chrom-on-Glass-Maske

2. Entwickeln der Lackschicht

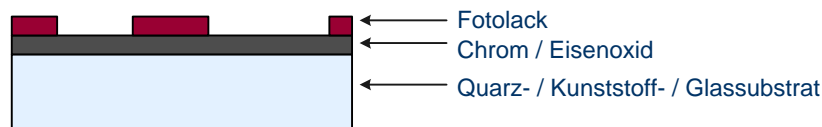


Abb. 6.13: Entwickeln der Lackschicht

3. Ätzen der Chromschicht

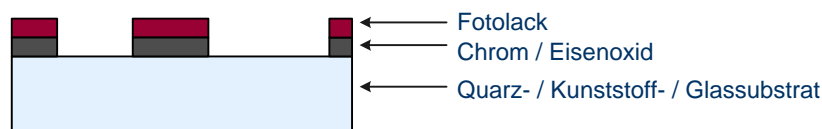


Abb. 6.14: Ätzen der Chromschicht

4. Lackentfernen



Abb. 6.15: Lackentfernen

5. Anbringen des Pellicles

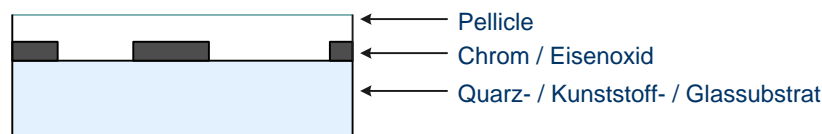


Abb. 6.16: Anbringen des Pellicles

6.5.3 Maskentypen

Neben der klassischen Chrommaske (COG, Chrome On Glass) gibt es noch weitere Maskentypen, die eine verbesserte Strukturauflösung ermöglichen. Hauptproblem der COG-Maske ist die Beugung, die das Licht an den Strukturkanten erfährt. Dadurch fällt das Licht nicht nur senkrecht auf den Wafer, sondern wird auch in den Schattenbereich abgelenkt, wo der Fotolack nicht belichtet werden soll.

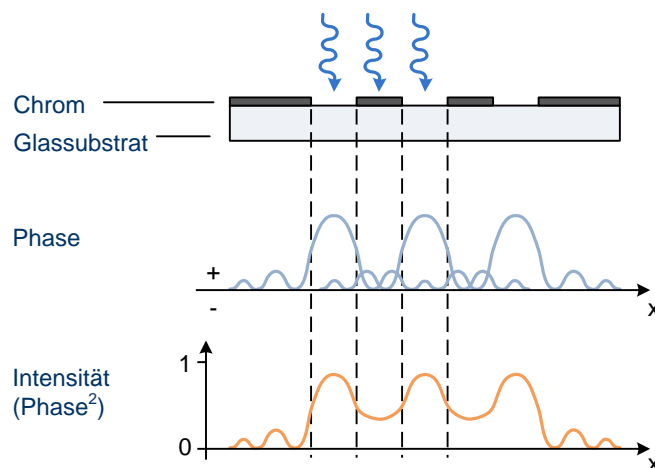


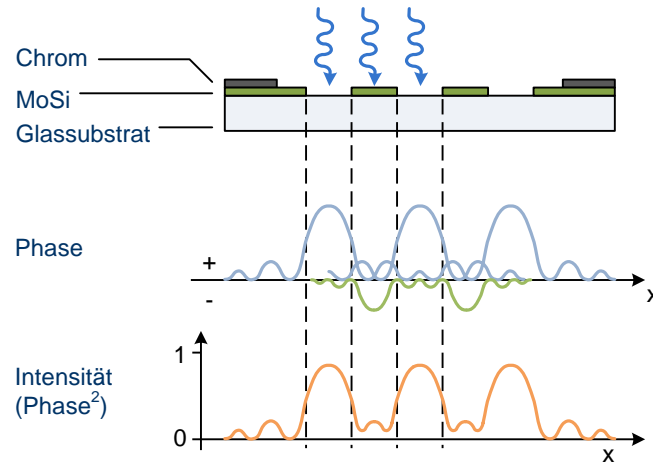
Abb. 6.17: Intensitätsprofil einer Chrom On Glass-Maske

Mit verschiedenen Maßnahmen versucht man nun, die Intensität des gebeugten Lichts zu vermindern. Diese werden im Folgenden anhand der unterschiedlichen Maskentypen näher beschrieben.

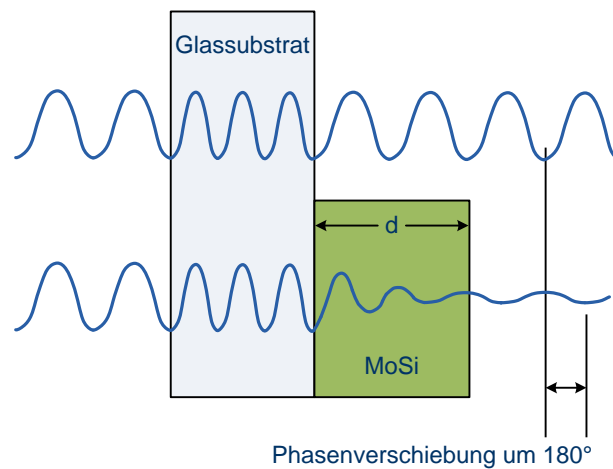
Attenuated Phase Shift Mask (AttPSM):

Bei der so genannten Halbtonmaske oder weichen Phasenmaske bildet eine Schicht aus Molybdänsilicid (MoSi) den strukturgebenden Teil, eine Chromschicht gibt es hier nicht. Die Dicke der MoSi-Schicht ist so gewählt, dass das Licht beim Durchgang eine Phasenverschiebung um 180° erfährt gegenüber dem Licht, das lediglich Glas durchläuft. Gleichzeitig ist die Schicht je nach Molybdänanteil im Silicium zu 6 % oder 18 %

lichtdurchlässig (bei einer Belichtungswellenlänge von 193 nm), das Licht wird also abgeschwächt (attenuate, engl.: vermindern). Die gegenläufigen Lichtwellen löschen sich so unterhalb der MoSi-Strukturen nahezu aus. Zusätzlich kann in Bereichen, die nicht zur Belichtung benötigt werden, Chrom aufgebracht werden, um Licht vollständig auszublenden. Diese Masken werden als Tritonemaske bezeichnet.



(a)



(b)

Abb. 6.18: (a) Intensitätsprofil einer Attenuated Phase Shift Mask, (b) Prinzip der Phasenschiebung mittels Molybdänsilicid

Chromfreie Phasenschiebermaske:

Die chromfreien Masken besitzen keine strukturgebende Beschichtung. Die Phasen-

verschiebung wird durch Gräben erzeugt, die direkt in die Glasplatte geätzt sind. Die Herstellung dieser Masken gestaltet sich daher schwierig, da der Ätzvorgang mitten im Glas gestoppt werden muss. Im Gegensatz zu Ätzprozessen, bei denen eine Schicht vollständig durchgeätzt wird und Änderungen im Plasma Auskunft darüber geben, wann die darunterliegende Schicht freigelegt ist, erhält man hier keine Information, wann die benötigte Tiefe erreicht ist.

Alternating Phase Shift Mask:

Bei der alternierenden Phasenmaske werden wie bei der chromfreien Maske Gräben direkt in das Glassubstrat geätzt, jedoch abwechselnd (alternierend) mit ungeätzten Bereichen. Zusätzlich werden Stellen mit Chrom beschichtet um die Lichtintensität an entsprechenden Stellen herabzusetzen.

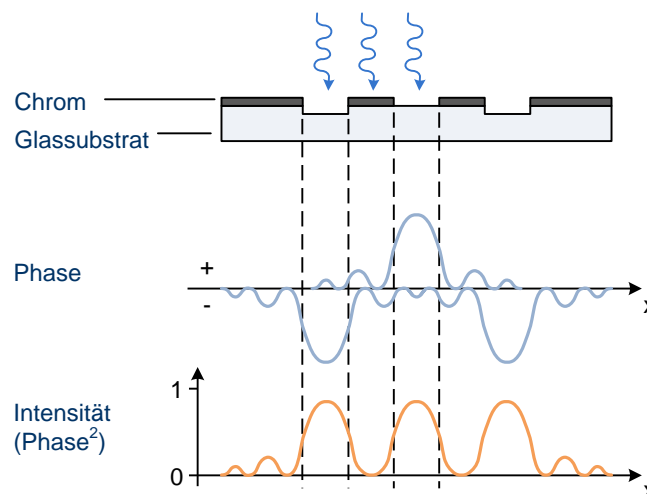


Abb. 6.19: Intensitätsprofil einer Alternating Phase Shift Mask

Dadurch ergeben sich jedoch Bereiche mit undefinierter Phasenverschiebung (Abb. 6.20), so dass bei diesem Maskentyp, der eine sehr hohe Auflösung ermöglicht, grundsätzlich zweimal belichtet werden muss. Die erste Maske enthält dabei die Strukturen die in x-Richtung verlaufen, die zweite Maske die Struktur in y-Richtung.

6.5.4 Belichtungsverfahren der nächsten Generation

Zukünftig muss eine Umstellung der heute eingesetzten Anlagen und der transparenten Fotomasken erfolgen. Extreme UV-Strahlung (EUV, Wellenlänge 13,5 nm), wie sie voraussichtlich bei der nächsten Generation von Belichtungsverfahren eingesetzt wird,

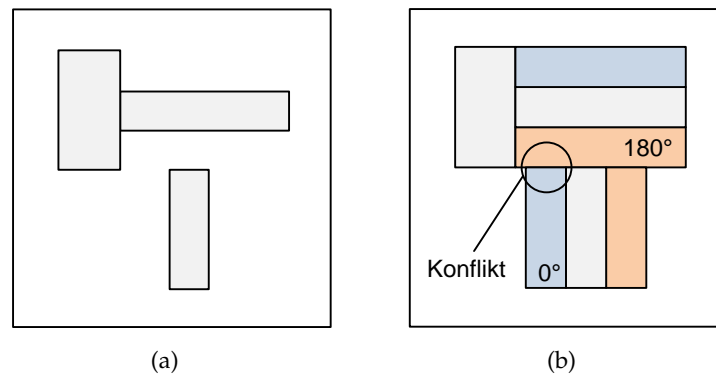


Abb. 6.20: Zielstruktur auf dem Wafer und entsprechende Maskenstruktur

wird in normaler Atmosphäre sowie vom Glassubstrat nahezu vollständig absorbiert. Aus diesem Grund müssen die Prozesse unter Vakuum stattfinden, zur Fokussierung müssen Spiegel eingesetzt werden. Die Masken werden dann anstelle von transparenten Bereichen spiegelnde Oberfläche besitzen.

Nicht zuletzt wegen des enormen technischen wie auch finanziellen Aufwands versucht man die Lithografie mit refraktiven Optiken (z.B. Linsen) und den herkömmlichen Quarzmasken so lange wie möglich am Leben zu erhalten. Belichtungsverfahren mit phasenschiebenden Fotomasken und Immersionstechniken haben die konventionelle Lithografie schon sehr weit vorangetrieben, so können Strukturen von 32 nm noch immer mit 193 nm Wellenlänge belichtet werden.

7 Nasschemie

7.1 Ätztechnik allgemein

In der Halbleitertechnologie müssen verschiedenste Schichten geätzt werden. Entweder zur ganzflächigen Entfernung oder zum Übertragen eines strukturierten Fotolackfilms in eine darunter liegende Schicht. Dabei lässt sich die Ätztechnik in nasschemisches Ätzen und Trockenätzen unterteilen. Man unterscheidet hierbei grundsätzlich zwischen isotropen und anisotropen Prozessen, sowie dem chemischen und dem physikalischen Charakter.

Bei einem isotropen Ätzprozess geschieht die Ätzung in alle Richtungen. So werden Schichten nicht nur in ihrer Dicke abgetragen, sondern auch in ihrer Ausdehnung in horizontaler Richtung. Bei anisotropen Ätzungen wird die Schicht nur senkrecht abgetragen. Je nach Prozessschritt kann ein isotroper oder ein anisotroper Ätzvorgang erwünscht sein.

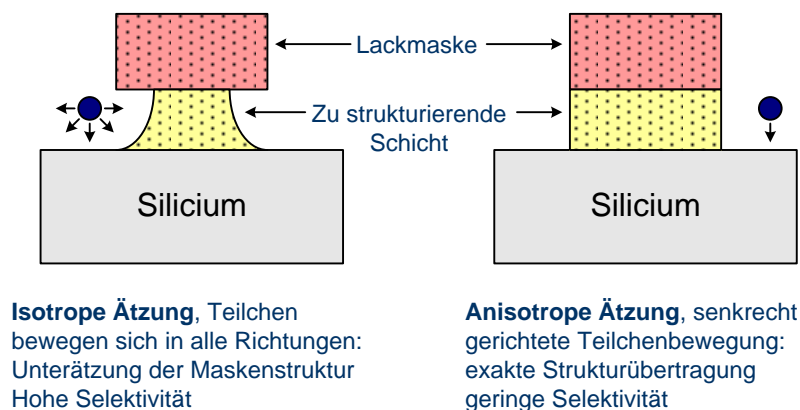


Abb. 7.1: Isotroper und anisotroper Ätzvorgang

Ein wichtiger Parameter der Ätzprozesse ist die Selektivität. Diese gibt das Verhältnis vom Abtrag der zu ätzenden Schicht zum Abtrag der jeweils anderen Schicht. Beträgt

die Selektivität 2:1, dann wird die zu strukturierende Schicht, z.B. Oxid, doppelt so schnell geätzt wie die Lackschicht.

Die Nasschemie wird aber nicht allein zum Ätzen von Schichten verwendet, sondern kommt auch bei diesen Prozessen zum Einsatz:

- **Nassätzen:** zum ganzflächigen Entfernen von dotierten und undotierten Oxidschichten
- **Scheibenreinigung**
- **Lackentfernen**
- **Rückseitenbehandlung:** Entfernen von Schichten die bei Ofenprozessen auf den Scheibenrückseiten entstanden sind
- **Polymerentfernung:** Entfernung von Polymeren die beim Plasmaätzen entstehen und sich auf den Scheiben festsetzen

Auf Grund des meist isotropen Ätzprofils wird das Nassätzen kaum zur Strukturierung verwendet.

7.2 Nassätzen

7.2.1 Prinzip

Das Prinzip beim Nassätzen ist die Umwandlung des festen Materials der Schicht in flüssige Verbindungen unter Zuhilfenahme einer chemischen Lösung. Die Selektivität ist sehr hoch, da die eingesetzten Chemikalien genau auf die vorhandenen Schichten abgestimmt werden können; sie beträgt bei den meisten Lösungen mehr als 100:1.

7.2.2 Anforderung

Folgende Anforderungen müssen von den chemischen Lösungen erfüllt werden:

- die Maskierschicht darf nicht angegriffen werden
- die Selektivität muss hoch sein
- die Ätzung muss durch Verdünnung mit Wasser gestoppt werden können

- es dürfen sich keine gasförmigen Reaktionsprodukte bilden, da die Bläschen vereinzelt Bereiche abschatten können
- konstante Ätzrate über lange Zeit
- die Reaktionsprodukte müssen direkt gelöst werden, damit keine Partikel die Ätzlösung verunreinigen
- gute Umweltverträglichkeit und leichte Entsorgung

7.2.3 Tauchätzung

Beim Tauchätzverfahren wird eine ganze Horde mit Wafern auf einmal in einem Becken mit der Ätzlösung prozessiert. Durch Filter und Umwälzpumpen werden Partikel von den Wafern ferngehalten. Da die Konzentration der Ätzlösung mit zunehmender Menge an bearbeiteten Scheiben abnimmt, muss sie oft erneuert werden.

Die Ätzrate, also der Materialabtrag pro Zeit, muss genau bekannt sein, um reproduzierbare Ätzungen durchführen zu können. Eine exakte Temperierung der Ätzlösung ist erforderlich, da die Ätzrate der meisten chemischen Lösungen mit der Temperatur zunimmt.

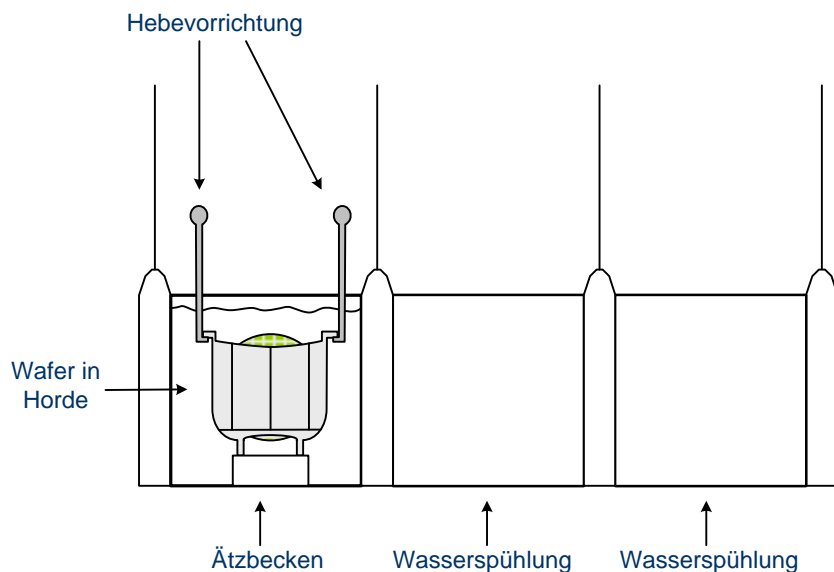


Abb. 7.2: Tauchätzung

Die Hebevorrichtung kann die Horde waagerecht und senkrecht transportieren. Nach-

dem die Wafer geätzt wurden wird die Ätzung auf den Wafern durch Spülen in mehreren Tauchbecken gestoppt; anschließend kommen die Wafer in eine Trockenschleuder.

Die Vorteile der Tauchätzung sind die hohe Durchsatzrate an Wafern und der relativ einfache Aufbau der Ätzanlagen. Die Homogenität des Schichtabtrags ist jedoch gering.

7.2.4 Sprühätzung

Die Sprühtechnik ist vergleichbar mit der Sprühtwicklung in der Fototechnik. Durch die Rotation auf dem Chuck unter stetiger Zugabe frischer Ätzlösung ist die Homogenität sehr gut. Bläschen können sich hier auf Grund der schnellen Drehung nicht bilden, jedoch ist auch hier der Nachteil, dass jeder Wafer einzeln prozessiert werden muss.

Alternativ zu der sequentiellen Prozessierung gibt es auch die Sprühätzung mehrerer Horden gleichzeitig in einer Trommel, bei der die Wafer schnell um die Sprühvorrichtung im Zentrum kreisen. Direkt im Anschluss werden die Scheiben unter Rotation in heißer Stickstoffatmosphäre getrocknet.

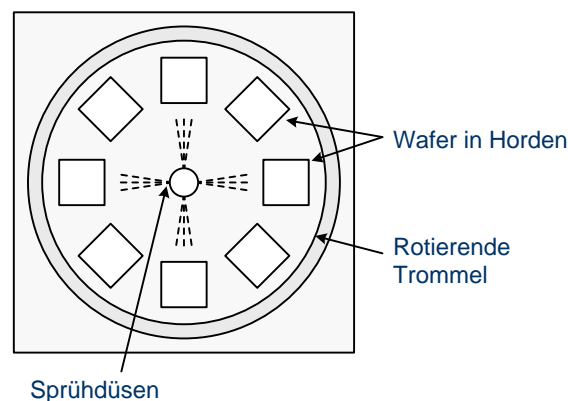
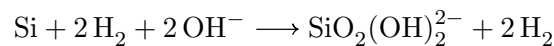


Abb. 7.3: Sprühätzung

7.2.5 Anisotrope Siliciumätzung

Obwohl in einer Ätzlösung die Teilchen in der Flüssigkeit die Schicht auf dem Wafer in jeder Richtung abätzen können gibt es auch Verfahren, mit denen man bei der Nassätzung ein nahezu anisotropes Ätzprofil erhält. Dabei nutzt man die unterschiedliche Ätzrate auf den verschiedenen Kristallstrukturen aus. (100)- und (110)-orientierte

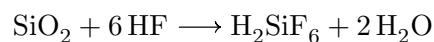
Kristallebenen lassen sich deutlich schneller abtragen als (111)-orientierte Kristalle. So lassen sich V-Gräben (100-Si) oder senkrechte Wände (110-Si) ätzen. Die Ätzung erfolgt mit Kali-, Natron- oder Lithiumlauge (KOH, NaOH, LiOH) oder mit einer EDP-Lösung (ein Gemisch aus Wasser, Pyrazin, Brenzkatechin und Ethylendiamin). Verantwortlich für die Reaktion ist in jedem Fall die OH^- -Gruppe (Hydroxidgruppe) der Stoffe:



Anisotrop wirkende Ätzlösungen sind für mikroelektronische Schaltungen jedoch uninteressant.

7.2.6 Ätzlösungen für isotrope Ätzungen

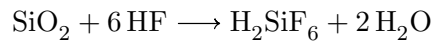
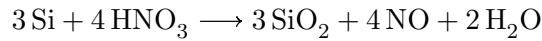
Für die jeweiligen Materialien stehen unterschiedliche Ätzlösungen zur Verfügung. Oxidschichten werden in der Halbleiterindustrie mit Flusssäure HF geätzt:



Die Lösung wird dabei mit NH_4F gepuffert um die HF-Konzentration konstant zu halten. Bei einer Mischung aus einer 40-prozentigen NH_4F -Lösung und 49-prozentigen Flusssäure (Verhältnis 10:1) beträgt die Ätzrate bei thermischem Oxid 50 nm/min. TEOS-Oxide werden mit ca. 150 nm/min und PECVD-Oxide mit ca. 350 nm/min deutlich schneller geätzt. Die Selektivität gegenüber kristallinem Silicium, Siliciumnitrid und Polysilicium ist wesentlich größer als 100:1.

Siliciumnitrid wird mit heißer Phosphorsäure (H_3PO_4) geätzt. Dabei ist die Selektivität zu SiO_2 jedoch mit 10:1 sehr gering. Bei Polysilicium wird die Selektivität zum Nitrid im Wesentlichen vom Gehalt der Phosphorsäure bestimmt.

Kristallines und polykristallines Silicium, werden zunächst mit Salpetersäure (HNO_3) oxidiert, das Siliciumdioxid wird dann mit Flusssäure geätzt:



Aluminium wird bei ca. 60 °C mit einer Mischung aus Salpeter- und Phosphorsäure geätzt, Titan mit einer Lösung aus Ammoniakwasser (NH₄OH), Wasserstoffperoxid (Titmus, H₂O₂) und Wasser (Verhältnis 1:3:5). Da die Lösung auch Silicium angreift wenn das Peroxid verbraucht ist, ist die Standzeit der Lösung gering.

Generell eignet sich das nasschemische Ätzen für das abtragen kompletter Schichten auf einem Wafer. Die Selektivität von der zu ätzenden Schicht zu der darunter liegenden ist meist sehr gut, so dass keine Gefahr besteht falsche Schichten abzutragen. Zudem ist der Abtrag pro Zeiteinheit sehr hoch, in Tauchätzverfahren lassen sich viele Scheiben gleichzeitig ätzen. Bei kleinen Strukturen kann die Nassätzung jedoch nicht eingesetzt werden, da das isotrope Ätzprofil dafür nicht geeignet ist. In diesem Fall werden die Schichten mit Trockenätzverfahren anisotrop abgetragen.

7.3 Scheibenreinigung

7.3.1 Der Reinraum

Die Halbleiterfertigung findet in Reinräumen statt um die hochkomplexen Schaltungen der Halbleiterbauteile vor Verunreinigungen zu schützen, die die Funktionsfähigkeit der Elemente beeinflussen können. Die Reinräume werden nach der Größe der Partikel und deren Anzahl pro Kubikfuß (= cbf; 1 Fuß = 30,48 cm) klassifiziert:

In einem Reinraum Klasse 100 ist die maximal zulässige Partikelgröße 0,5 µm, es dürfen nicht mehr als 100 solcher Partikel in einem Kubikfuß Luft vorhanden sein. In einem Operationssaal beträgt die Reinraumklasse 100 oder 1000. Stadtluft hat in einem Volumen von 1 cbf 15-100 Millionen Partikel der Größe 5 µm.

Die Luft in Reinräumen für die Mikroelektronik wird über Feinstfilter gereinigt und dann durch die Decke geblasen. Durch Löcher im Boden wird die Luft abgesaugt. Dieser Laminarstrom der Luft transportiert Partikel von oben nach unten durch den Boden ab, damit keine Verunreinigungen von außen in den Reinraum eindringen können

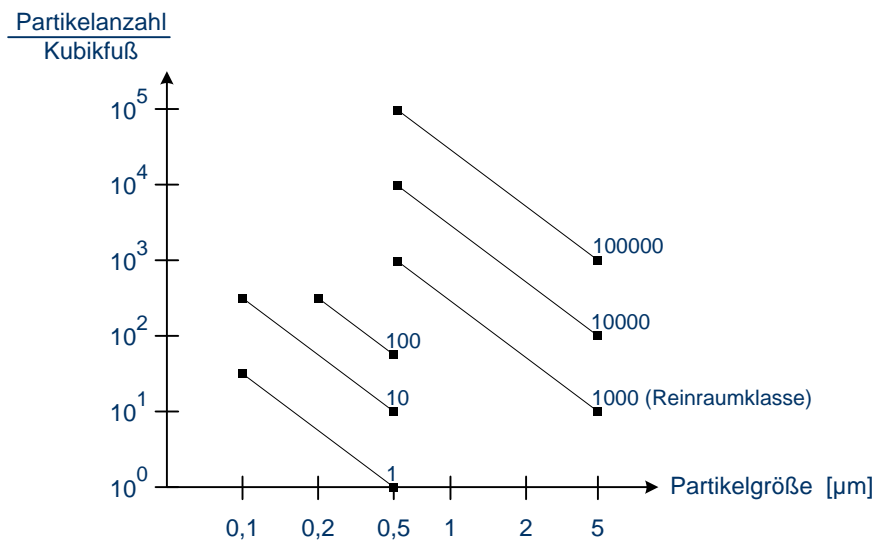


Abb. 7.4: Reinraumklassen

herrscht dort immer etwas Überdruck. Da die von den Produktionsanlagen erwärmte Luft nach oben steigt und sich so mit dem nach unten gerichteten Laminarstrom Verwirbelungen ergeben können, kann der Laminarstrom auch von unten nach oben verlaufen. Dies ist vor allem dann der Fall, wenn die Wafer in Boxen transportiert werden, die direkt an den Anlagen andocken und so geöffnet werden, dass Partikel aus der Umgebung nicht zu den Wafern gelangen können. Dabei handelt es sich in der Regel um sogenannte FOUPs (Front Opening Unified Pod).

Die Fertigung im Reinraum ist in der Abbildung nur der Bereich zwischen rosafarbener Lüftungsanlage und dem gelben Bodenbereich. Darunter befindet sich das Basement mit den Versorgungsanlagen (Pumpen etc.), oben der gesamte Lüftungsbereich. Der Reinraum ist meistens in einzelne Gänge unterteilt (im Bild: Mittelgang, T8, M23), um die Luftverwirbelungen möglichst gering zu halten. Viele Reinträume sind von einem so genannten Grauraum umgeben, in dem sich die Anlagen befinden. Über eine Wand getrennt befinden sich dann nur die Bedienfelder und die Schleusen zum Einbringen der Wafer in die Anlagen im Reinraum.

Das Personal trägt spezielle Reinraumanzüge, die keine Partikel absondern. Dabei bedeckt der Anzug, je nach Bedarf den kompletten Körper. Der Kopf wird entweder mit einem simplen Haarnetz bzw. einer Haube und einem Mundschutz oder mit einer kompletten Maske bedeckt. Zusätzlich gibt es besonders partikelarme Schuhe, evtl. Unterbekleidung und Handschuhe. Im Eingangsbereich des Reinraums können sich

eine oder mehrere Luftduschen befinden, die vor Betreten noch einmal alle Partikel vom Anzug abblasen.

7.3.2 Arten der Verunreinigung

Trotz des Reinraums gibt es verschiedene Arten von Verunreinigungen, die hauptsächlich vom Personal in der Fertigungslinie, der Umgebungsluft, von Chemikalien (Gasen, Lösungen) und den Anlagen verursacht werden:

- **mikroskopische Verunreinigungen:** z.B. Partikel aus der Umgebungsluft oder Gasen
- **molekulare Verunreinigungen:** z.B. Kohlenwasserstoff aus Öl in den Pumpsystemen
- **ionische Verunreinigung:** z.B. Handschweiß
- **atomare Verunreinigung:** z.B. Schwermetalle aus Lösungen, Abrieb von Festkörpern

7.3.3 Mikroskopische Verunreinigungen

Mikroskopische Verunreinigungen sind Partikel, die sich an der Waferoberfläche anlagern. Quellen dieser Verunreinigungen sind die Umgebungsluft, Kleidung des Personals, Abriebe an beweglichen Teilen in Prozessanlagen, unzureichend gefilterte Flüssigkeiten, wie Ätzhemikalien oder Reinigungs- und Entwicklerlösungen, oder Ätzrückstände nach Prozessen beim Trockenätzen.

Mikroskopische Verunreinigungen verursachen Abschattungseffekte, wenn sie z.B. an der Scheibenoberfläche die Belichtung des Fotolacks verhindern, bei der Kontaktbelichtung kommt es zu schlechten Auflösungen, wenn sich relativ große Partikel zwischen Maske und Scheibe befinden. Des Weiteren können sie bei Implantationsprozessen oder beim Trockenätzen beschleunigte Ionen von der Scheibenoberfläche fernhalten.

Partikel können auch von Schichten eingeschlossen werden, so dass Unebenheiten entstehen. Nachfolgende Schichten können an diesen Stellen aufplatzen oder es sammelt sich Fotolack an, in Folge dessen es durch die zu dicke Lackschicht zu nicht vollständig belichteten Bereichen kommt.

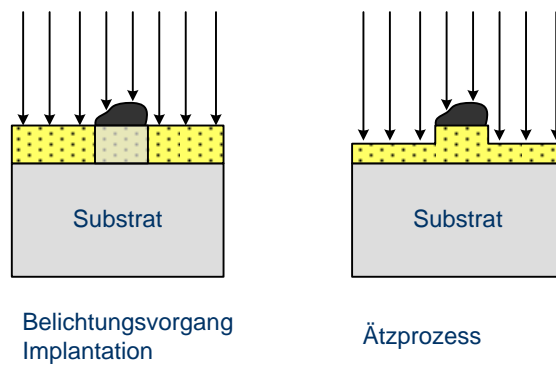


Abb. 7.5: Abschattung bei Ätz- und Implantationsprozessen durch Partikel

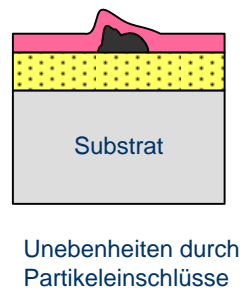


Abb. 7.6: Partikeleinschluss

7.3.4 Molekulare Verunreinigungen

Molekulare Verunreinigungen resultieren aus Lack- und Lösemittelresten auf den Wafern oder durch Ölnebelablagerungen aus Vakuumpumpen. Diese Verunreinigungen können sich sowohl auf der Scheibenoberfläche befinden, als auch in die Schichten eindiffundieren. Die oberflächlich anhaftenden Verunreinigungen behindern teilweise die Haftung nachfolgender Schichten, im Falle von Metallisierungen (dünne Leiterbahnen) erheblich. Durch die Diffusion in Oxide wird die elektrische Belastbarkeit dieser Schichten verschlechtert.

7.3.5 Alkalische und metallische Verunreinigungen

Die Hauptquelle dieser Verunreinigungen ist der Mensch, der ständig über die Haut, aber auch über die Atemluft Salze aussondert. Aber auch durch unzureichend deionisiertes Wasser gelangen (Alkali-)Ionen von Natrium oder Kalium auf die Scheiben. Schwermetalle die in Ätzlösungen vorhanden sind können zur Kontamination führen.

Durch Strahlen in Anlagen (Implanter, Trockenätzen) kann Material von Wänden abgesputtert werden, das sich dann auf den Scheiben niederschlägt.

Ionische Verunreinigungen beeinflussen beispielsweise die elektrischen Eigenschaften von MOS-Transistoren, da ihre Ladung die Schwellspannung - also die Spannung ab der der Transistor leitend wird - verändert. Schwermetalle wie Eisen oder Kupfer liefern freie Elektronen, so dass die Leistungsaufnahme in Dioden steigt. Metalle können aber auch Rekombinationszentren für freie Ladungsträger bilden, so stehen in der Schaltung nicht genug freie Ladungsträger zur einwandfreien Funktion zu Verfügung.

7.3.6 Reinigungstechniken

Die Scheiben werden nach jedem nasschemischen Prozessschritt in Reinstwasser gespült, aber auch nach anderen Prozessen werden die Scheiben von Verunreinigungen befreit. Dabei gibt es verschiedene Reinigungstechniken, die die unterschiedlichen Verschmutzungen beseitigen. Der Verbrauch an Reinstwasser in einer Halbleiterfertigung ist dabei extrem hoch und beträgt mehrere Millionen Liter im Jahr. Dabei befinden sich im Reinstwasser fast keine Verunreinigungen mehr, 1-2 ppm (parts per million = Anzahl an Verunreinigungen pro Millionen Wasserteilchen) sind erlaubt (Beispiel, Stand 2002). Das Wasser wird meist direkt vor Ort in Aufbereitungsanlagen gereinigt.

Eine Möglichkeit der Scheibenreinigung ist das Ultraschallbad, bei dem die Wafer in eine Lösung aus Wasser und Ultraschallreinigungs- und Netzmitteln gegeben werden. Durch die Ultraschallanregung lösen sich Partikel von der Oberfläche, Metalle und molekulare Verunreinigungen werden vom Reinigungsmittel teilweise gebunden. Nicht stark anhaftende Partikel können auch mit Stickstoff abgeblasen werden.

Zum Entfernen organischer Verunreinigungen wie Fett, Öl oder Hautschuppen eignen sich Lösungsmittel wie Aceton oder Ethanol. Diese können jedoch Kohlenstoffrückstände hinterlassen.

Ionische Verunreinigungen (Ionen von Kalium, Natrium etc.) werden durch Spülen mit deionisiertem Wasser entfernt. Auch die Reinigung mit rotierenden Bürsten und einer Reinigungsflüssigkeit ist möglich. Jedoch werden bei strukturierten Scheiben die Partikel an den Kanten angelagert, die Bürsten können die Scheibenoberfläche beschädigen. Kontaktlöcher oder andere Vertiefungen können mit einer Hochdruckreinigung bei ca. 50 bar ausgespült werden. Hierbei werden jedoch keine ionischen oder metallischen Verunreinigungen entfernt.

Die Reinigung mit Wasser und verschiedenen Reinigungsmitteln reicht jedoch oft nicht aus. Oftmals müssen Verunreinigungen mit aggressiven Ätzlösungen entfernt oder Oberflächen gezielt minimal abgetragen werden. Gemische aus Wasserstoffperoxid und Schwefelsäure oder Ammoniak oder die Carosche Säure/Piranha-Lösung (Schwefelsäure mit Wasserstoffperoxid) können organische Verunreinigungen bei ca. 90 °C durch Oxidation ablösen.

Ein Gemisch aus Salzsäure und Wasserstoffperoxid bildet Alkalimetalle zu leicht löslichen Chloriden (Salzen) aus, Schwermetalle bilden Komplexe und gehen in Lösung. Mit Flusssäure kann natürliches Oxid entfernt werden, mit einer gezielt aufgetragenen Schicht Oxid mit Wasserstoffperoxid lässt sich die Oberfläche dagegen auch schützen.

Neben der Spülung der Scheiben nach jedem nasschemischen Prozessschritt, durchlaufen die Wafer bei einer vollständigen Reinigung eine Reihe von Reinigungsschritten nacheinander. Dabei ist die Reihenfolge der Reinigungsverfahren wichtig, da sich diese teilweise gegenseitig in der Reinigungswirkung behindern. Die Reinigungssequenz sieht beispielsweise wie folgt aus:

- Abblasen von Partikeln mit Stickstoff
- Reinigung im Ultraschallbad
- Entfernen von organischen Verunreinigungen mit H_2SO_4 - H_2O_2 (Carosche Säure)
- Entfernen feinerer organischer Verunreinigungen mit einer NH_4 - H_2O_2 -Lösung
- Entfernen von metallischen Verunreinigungen mit Salzsäure und Wasserstoffperoxid
- Trocknen der Scheiben in der Trockenschleuder unter heißer Stickstoffatmosphäre

Nach jedem Reinigungsschritt werden die Scheiben in Reinstwasser gespült, gegebenenfalls natürliches Oxid mit Flusssäure entfernt. Je nach aktueller Scheibenoberfläche sehen die Reinigungssequenzen unterschiedlich aus, da die Lösungen manche Schichten angreifen. Bei den immer kleiner werdenden Strukturen wird auch die Reinigung mehr und mehr erschwert. Nicht nur, dass sich kleine Öffnungen nicht so leicht erreichen lassen. Die Oberflächenspannung und Kapillarwirkungen können Strukturen umkippen lassen und so die Schaltung zerstören.

8 Trockenätzen

8.1 Übersicht

8.1.1 Allgemein

Bei der nasschemischen Ätzung werden Schichten in der Regel isotrop abgetragen, doch gerade bei kleinen Strukturen ist ein anisotropes Ätzprofil wichtig. Dafür eignen sich die Trockenätzverfahren, die eine ausreichende Selektivität bieten. Die Verfahren ermöglichen reproduzierbare homogene Ätzungen sämtlicher Schichten, die in der Halbleiterfertigung zum Einsatz kommen. Dabei können neben anisotropen Ätzprofilen auch isotrope realisiert werden. Trotz der hohen Anlagenkosten und der sequentiellen Einzelscheibenbearbeitung hat sich das Trockenätzen gegen die nasschemische Ätzung durchgesetzt.

8.1.2 Wichtige Größen beim Trockenätzen

Ätzrate r : Die Ätzrate beschreibt den Ätzabtrag pro Zeit und wird i. d. R. in Nanometer pro Minute angegeben.

$$r = \frac{\text{Ätzabtrag } \Delta z}{\text{Zeit } \Delta t}$$

Anisotropiefaktor f : Der Anisotropiefaktor beschreibt das Verhältnis von horizontaler Ätzrate r_h zu vertikaler Ätzrate r_v .

$$f = 1 - \frac{r_h}{r_v}$$

Zur Strukturübertragung sind stark anisotrope Prozesse erwünscht, also Ätzungen nur in vertikaler Richtung, so dass die Lackmaske nicht unterätzt wird. Für anisotrope

Ätzungen ergibt sich $f \rightarrow 1$, und dementsprechend für isotrope Prozesse $f \rightarrow 0$.

Selektivität S_{jk} zwischen Material j und Material k: die Selektivität beschreibt das Verhältnis der Ätzrate zweier Materialien, z. B. von zu strukturierender Schicht (j) und Lackmaske (k).

$$S_{jk} = \frac{r_j}{r_k}$$

Ob eine hohe oder geringe Selektivität gewünscht ist, hängt vom jeweiligen Prozess ab. Bei der Strukturierung von Schichten sollte der Wert möglichst groß sein, d. h. die zu strukturierende Schicht wird schneller abgetragen als die Lackmaske. Beim Re-flowrückätzen muss die Selektivität 1 betragen um Topografien gleichmäßig einzuebnen.

8.1.3 Trockenätzverfahren

Bei den Trockenätzverfahren werden Gase durch hochfrequente Wechselfelder angeregt, typisch sind 13,56 MHz und 2,45 GHz. Bei einem Druck zwischen 1 Pa und 100 Pa liegt die freie Weglänge der Teilchen bei wenigen Millimetern bis Zentimetern.

Es gibt generell drei verschiedene Arten des Trockenätzens:

- **Physikalisches Trockenätzen:** physikalischer Abtrag der Waferoberfläche durch beschleunigte Teilchen
- **Chemisches Trockenätzen:** Reaktion zwischen Gas und Scheibenoberfläche
- **Chemisch-Physikalisches Trockenätzen:** physikalisches Ätzen mit chemischen Anteil

8.2 Trockenätzverfahren

8.2.1 Ionenstrahlätzen

Das Ionenstrahlätzen ist ein rein physikalisches Trockenätzverfahren. Dabei werden Argonionen als gerichteter Ionenstrahl mit 1-3 keV auf die Scheibe gelenkt. Durch die Energie der Teilchen schlagen diese Material aus der Oberfläche heraus.

Der Wafer wird dabei senkrecht oder gekippt zum Ionenstrahl in der Prozesskammer gehalten, die Ätzung erfolgt völlig anisotrop. Die Selektivität ist nur gering, da keine Unterscheidung der Materialien von den beschleunigten Ionen ausgeht. Das Gas und herausgeschlagenes Material werden vom Pumpensystem abgesaugt, jedoch setzen sich auch Partikel an den Kammerwänden und an vertikalen Kanten der Scheibe selbst ab, da das abgetragene Material nicht in den gasförmigen Zustand übergeht.

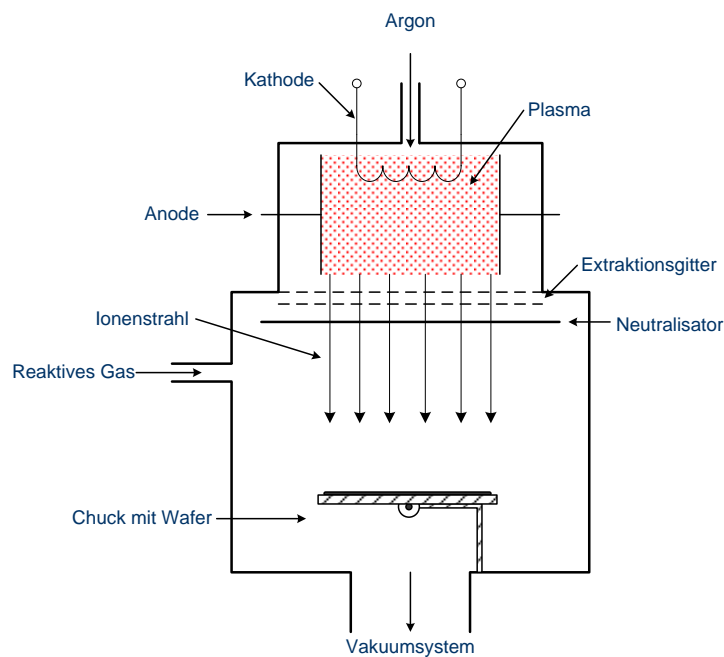


Abb. 8.1: Darstellung eines Ionenstrahl-Ätzreaktors

Um die Partikelablagerungen zu verhindern wird neben Argon ein reaktives Gas in die Kammer eingeleitet. Das Gas reagiert dabei mit den Argonionen und führt zu einem physikalischen Ätzprozess mit chemischem Charakter. Das Gas reagiert nun teilweise mit der Oberfläche, aber auch mit dem durch den physikalischen Anteil herausgeschlagenen Material, zu einem gasförmigen Reaktionsprodukt.

Mit diesem Verfahren können nahezu alle Materialien geätzt werden. Durch die senkrechte Bestrahlung ist der Abtrag an senkrechten Kanten sehr gering (hohe Anisotropie). Auf Grund der geringen Selektivität und der geringen Ätzrate (geringe Iondichte im Strahl) wird dieses Verfahren in der Halbleiterfertigung kaum mehr angewandt.

8.2.2 Plasmaätzen

Das Plasmaätzen ist ein rein chemisches Ätzen. Der Vorteil dabei ist, dass die Scheibenoberfläche nicht durch beschleunigte Ionen geschädigt wird. Das Ätzprofil ist auf Grund der frei beweglichen Gasteilchen isotrop, weshalb das Plasmaätzen meist zum Abtragen kompletter Schichten bei hoher Ätzrate eingesetzt wird.

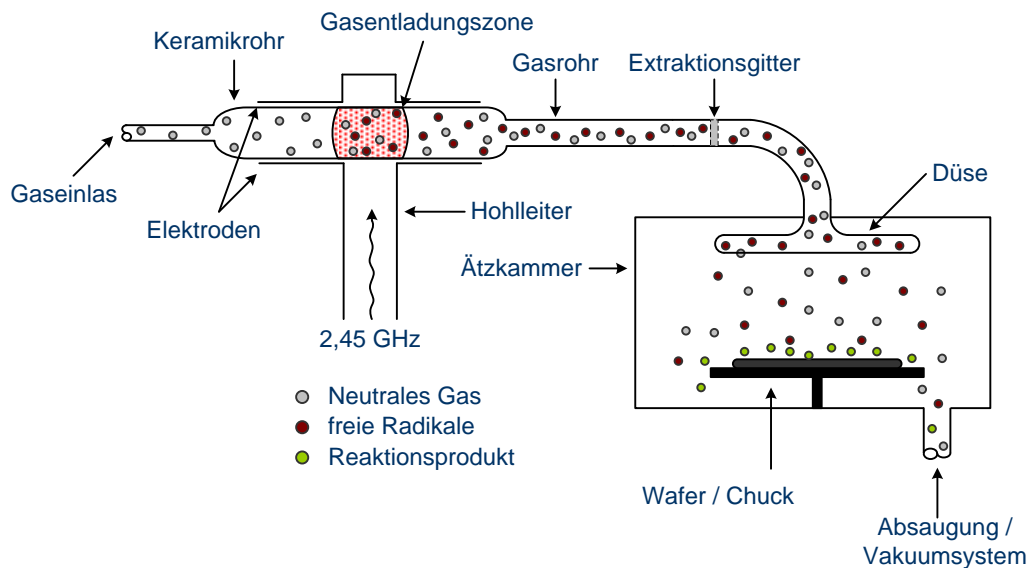
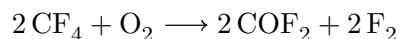


Abb. 8.2: Darstellung einer CDE-Ätzanlage

Eine Anlagenart beim Plasmaätzen ist der Down-Stream-Reaktor (Chemical Dry Etching, CDE). Dabei wird durch Anlegen einer hochfrequenten Spannung (typisch sind 2,45 GHz) ein Plasma durch Stoßionisation erzeugt. Der Raum der Gasentladung ist dabei vom Wafer getrennt, das Gas gelangt über ein Keramikrohr zum Wafer.

In der Zone der Gasentladung entstehen durch die Stöße der Gasmoleküle untereinander verschiedene Teilchen, unter anderem Radikale. Radikale sind Moleküle mit aufgespaltener Bindung, die durch ein ungepaartes Außenelektron sehr reaktionsfreudig sind. Als Neutrales Gas wird beispielsweise Kohlenstofftetrafluorid CF_4 eingeleitet und durch die Wechsellspannung in CF_2 und ein Fluormolekül F_2 zerlegt. Ebenso kann durch Zugabe von Sauerstoff Fluor von CF_4 abgespalten werden:



Das Fluormolekül kann nun durch die Energie in der Gasentladungszone in zwei einzelne Fluorradikale aufgespaltet werden: Fluorradikale (Radikale sind ungeladene Teilchen, ein einzelnes Fluoratom besitzt neun Protonen im Kern und neun Elektronen in der Atomhülle).

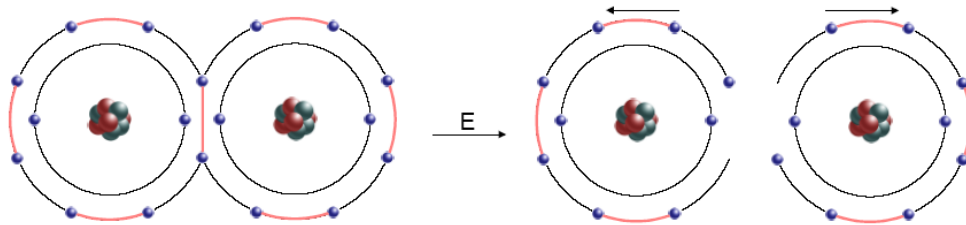


Abb. 8.3: Bildung von Radikalen

Neben den neutralen Radikalen entstehen weitere, teils geladene Teilchen (CF_4^+ , CF_3^+ , CF_2^+ , ...). Die Teilchen werden nun über das Gasrohr zur Ätzkammer geleitet. Geladene Teilchen werden entweder durch ein Extraktionsgitter abgefangen oder rekombinieren auf dem Weg wieder zu ungeladenen Molekülen. Auch die Fluorradikale werden teilweise wieder gebunden. Es gelangen jedoch genügend Radikale in die Ätzkammer, und reagieren dort mit der Scheibenoberfläche. Neutrale Teilchen nehmen nicht an der Reaktion teil und werden, so wie auch die entstandenen Reaktionsprodukte, abgesaugt.

Beispiele für Schichten die im CDE-Verfahren geätzt werden:

- **Silicium:** $\text{Si} + 4 \text{F} \longrightarrow \text{SiF}_4$
- **Siliciumdioxid:** $\text{SiO}_2 + 4 \text{F} \longrightarrow \text{SiF}_4 + \text{O}_2$
- **Silicium-Nitrid:** $\text{Si}_3\text{N}_4 + 12 \text{F} \longrightarrow 3 \text{SiF}_4 + 2 \text{N}_2$

8.2.3 Reaktives Ionenätzen

Die Ätzcharakteristik - Selektivität, Ätzprofil, Ätzrate, Homogenität, Reproduzierbarkeit - ist beim reaktiven Ionenätzen (Reactive Ion Etching, kurz RIE) durch die eingesetzten Gase und die Prozessparameter (Generatorleistung, Druck, Plattenabstand, Gasfluss) exakt einstellbar. Dabei ist sowohl ein isotropes als auch ein anisotropes Ätzprofil möglich. Damit ist das RIE-Ätzen, ein chemisch-physikalisches Ätzverfahren,

das wichtigste Ätzverfahren zur Strukturierung diverser Schichten in der Halbleiterfertigung.

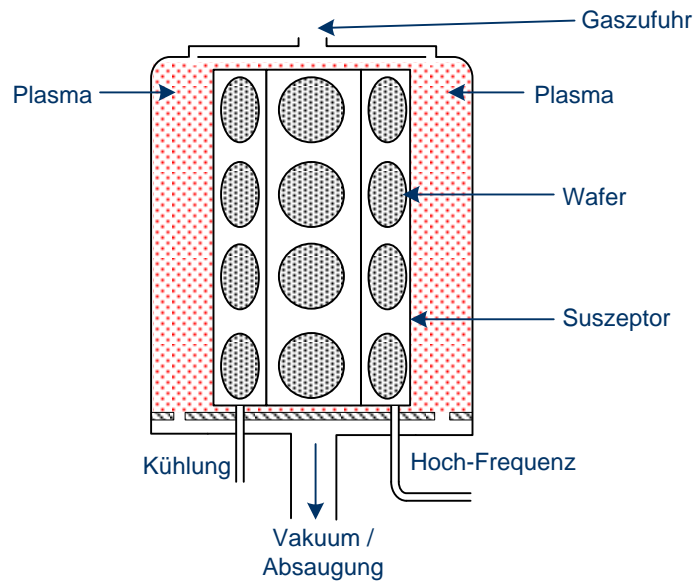


Abb. 8.4: Darstellung eines RIE-Reaktors in Hexodenbauform

In der Prozesskammer befinden sich die Wafer auf einer mit Wechselspannung gespeisten Elektrode (HF-Elektrode). Durch Stoßionisation wird ein Plasma erzeugt, das freie Elektronen und positiv geladene Ionen enthält. Bei positiver Spannung an der HF-Elektrode lagern sich die Elektronen an dieser an, und können sie auf Grund der Austrittsarbeit bei der positiven Halbwelle nicht verlassen, die Elektrode lädt sich somit auf bis zu 1000 V negativ auf (BIAS-Spannung). Die langsamen Ionen, die der schnellen Wechselspannung nicht folgen konnten, bewegen sich nun in Richtung der negativ aufgeladenen Elektrode mit den Wafern.

Ist die freie Weglänge der Ionen groß, treffen die Teilchen auf Grund ihrer Geschwindigkeit nahezu senkrecht auf die Scheiben. Dabei wird Material durch die beschleunigten Ionen von der Oberfläche herausgelöst (physikalisches Ätzen), teilweise reagieren die Teilchen aber auch chemisch mit dem Substrat. Vertikale Kanten werden dabei nicht getroffen, so dass hier kein Abtrag erfolgt und das Ätzprofil anisotrop ist. Die Selektivität ist durch den physikalischen Abtrag nicht sehr hoch, zudem wird die Scheibenoberfläche durch die beschleunigten Ionen geschädigt. Diese muss später durch thermische Behandlung ausgeheilt werden.

Der chemische Anteil der Ätzung geschieht durch die Reaktion von freien Radika-

len auf der Scheibenoberfläche und mit dem physikalisch abgetragenen Material, so dass sich dieses nicht wie beim Ionenstrahlätzen an den Kammerwänden oder auf den Scheiben anlagern kann. Durch Druckerhöhung nimmt die freie Weglänge der Teilchen ab, es passieren auf dem Weg zur Scheibe also viele Stöße zwischen den Teilchen, die so ihre Richtung fortlaufend ändern. So geschieht keine gerichtete Ätzung mehr auf der Scheibenoberfläche, der Ätzprozess nimmt einen stärker chemischen Charakter an, das Ätzprofil ist isotrop, die Selektivität erhöht sich.

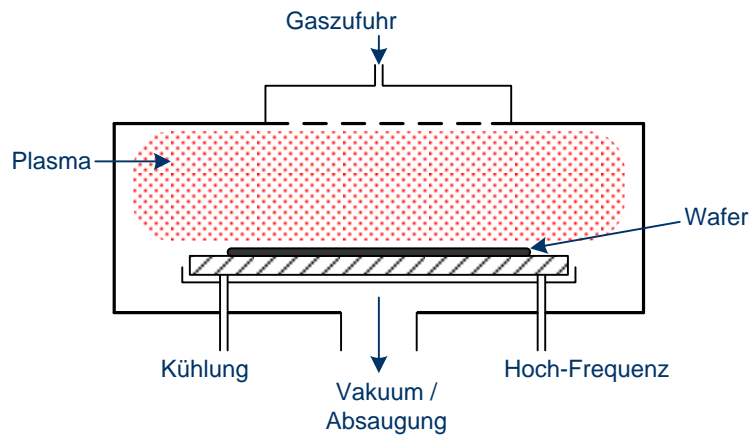


Abb. 8.5: Darstellung eines RIE-Reaktors in Parallelplattenbauweise

Durch eine Passivierung der Seitenwände bei einer Siliciumätzung wird ein anisotropes Ätzprofil erzielt. Dabei reagiert Sauerstoff in der Prozesskammer mit herausgelöstem Silicium der Waferoberfläche zu Siliciumdioxid, das an den vertikalen Kanten aufwächst. Eine Oxidschicht auf horizontalen Flächen wird durch den Ionenbeschuss entfernt, so dass eine Ätzung in die Tiefe fortschreiten kann. Die Ätzrate hängt in jedem Fall vom Druck, der Leistung des HF-Generators, den Prozessgasen, dem Gasdurchfluss und der Wafer- bzw. Elektroden temperatur ab.

Die Anisotropie nimmt bei steigender HF-Leistung, sinkendem Druck und abnehmender Temperatur zu. Die Homogenität der Ätzung wird von den Gasen, dem Elektrodenabstand und dem Elektrodenmaterial bestimmt. Bei zu geringem Elektrodenabstand ist das Plasma nicht gleichmäßig in der Kammer verteilt und führt so zur Inhomogenität. Bei größer werdendem Abstand nimmt die Ätzrate ab, da das Plasma auf ein größeres Volumen verteilt ist. Als Elektrodenmaterial hat sich Kohlenstoff bewährt. Da die Fluor- und Chlorgase auch Kohlenstoff abtragen, bewirkt die Elektrode eine gleichmäßige Belastung des Plasmas. Scheibenränder werden so nicht stärker belastet als die Scheibenmitte.

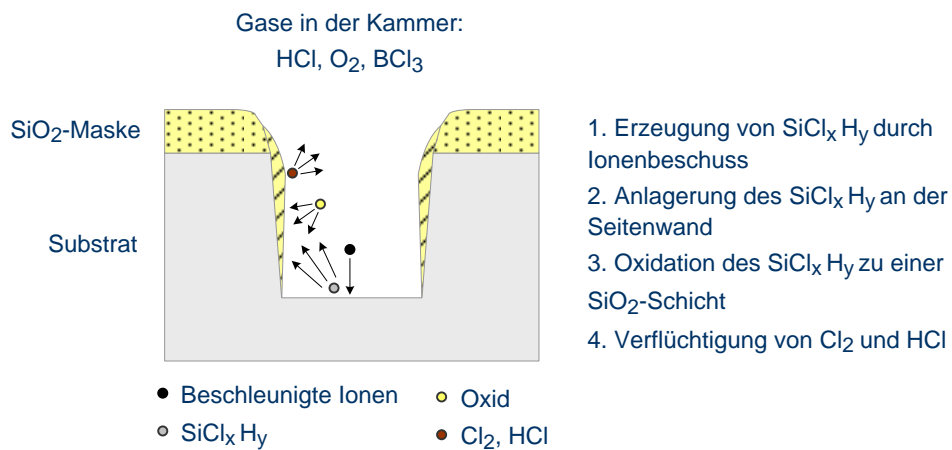


Abb. 8.6: Seitenwandpassivierung beim RIE-Ätzen

Die Selektivität und Ätzrate lassen sich sehr stark durch die eingesetzten Ätzgase beeinflussen. Bei Silicium und Siliciumverbindungen kommen vor allem Chlor und Fluor zum Einsatz.

Die Prozesse müssen dabei nicht nur mit einer Gasmischung ablaufen. Natürliches Oxid auf Polysilicium kann beispielsweise erst mit einer hohen Ätzrate und geringer Selektivität abgetragen werden. Das Polysilicium wird dann mit hoher Selektivität zur darunter liegenden Schicht geätzt.

Tab. 8.1: Übersicht über Prozessgase zum Trockenätzen verschiedener Schichten

Schicht	Prozessgase	Bemerkung
SiO ₂ , Si ₃ N ₄	CF ₄ , O ₂	F ätzt Si, O ₂ entfernt Kohlenstoff (C)
	CHF ₃ , O ₂	CHF ₃ wirkt als Polymer, bessere Selektivität zu Si
	CHF ₃ , CF ₄	
	CH ₃ F	verbesserte Selektivität von Si ₂ N ₄ über SiO ₂
	C ₂ F ₆ / SF ₆	
	C ₃ F ₈	erhöhte Ätzrate gegenüber CF ₄
Poly-Si	BCl ₃ , Cl ₂	keine Kontamination durch Kohlenstoff (C)
	SiCl ₄ , Cl ₂	
	HCl, O ₂	
	SiCl ₄ , HCl	
	O ₂ / SiCl ₄ , HCl	
	HBr / Cl ₂ / O ₂	verbesserte Selektivität zu Fotolack und SiO ₂
	SF ₆	hohe Ätzrate, gute Selektivität gegen SiO ₂
	NF ₃	hohe Ätzrate, isotrop
	HBr, Cl ₂	
monokrist. Silicium	HBr, NF ₃ , O ₂ / CF ₃ Br	höhere Selektivität gegen SiO ₂
	BCl ₃ , Cl ₂ / HBr, NF ₃	
Aluminium- Legierungen	Cl ₂	isotrope Ätzung
	BCl ₃	nur geringe Ätzrate
	BCl ₃ / Cl ₂ / CF ₄	anisotrope Ätzung
	BCl ₃ / Cl ₂ / CHF ₃	verbesserte Seitenwandpassivierung
	BCl ₃ / Cl ₂ / N ₂	höhere Ätzrate, keine Kohlenstoffkontamination

