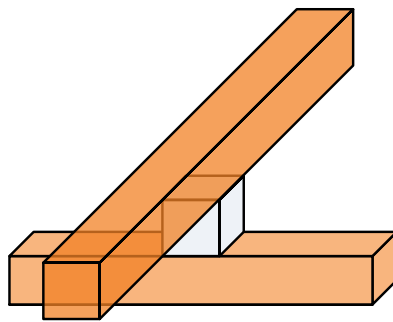


Halbleitertechnologie

von A bis Z



Metallisierung

www.halbleiter.org

Inhaltsverzeichnis

Abbildungsverzeichnis	II
Tabellenverzeichnis	III
1 Metallisierung	1
1.1 Anforderungen an die Metallisierung	1
1.2 Aluminiumtechnologie	2
1.2.1 Aluminium und Aluminiumlegierungen	2
1.2.2 Siliciumdiffusion	2
1.2.3 Elektromigration	4
1.2.4 Hillockwachstum	4
1.3 Kupfertechnologie	5
1.3.1 Kupfertechnologie	5
1.3.2 Damascene-Verfahren	6
1.3.3 Low-k-Technologie	10
1.4 Der Metall-Halbleiter-Kontakt	13
1.4.1 Kontaktierung von dotierten Halbleitern	13
1.4.2 n-Halbleiter-Kontakt	14
1.4.3 p-Halbleiter-Kontakt	17
1.4.4 Bändermodell eines p-n-Übergangs	17
1.5 Mehrlagenverdrahtung	18
1.5.1 Mehrlagenverdrahtung	18
1.5.2 BPSG-Reflow	18
1.5.3 Reflowrückätzen	19
1.5.4 Chemisch-Mechanisches Polieren	19
1.5.5 Kontaktierung der Metallisierungsebene	21

Abbildungsverzeichnis

1.1	Spikebildung	2
1.2	Kontakimplantation	3
1.3	Barrierschicht zwischen Aluminium und Substrat	3
1.4	Kantenverrundung zur konformen Aluminiumabscheidung	4
1.5	Prinzipieller Damascene-Schichtaufbau vor der Strukturierung	7
1.6	Grabenätzung und VIA-Lackmaske	7
1.7	Finale Struktur nach dem Ätzen und der Kupferabscheidung	8
1.8	Damascene-Struktur nach Metallisierung und Planarisierung	8
1.9	Damascene-Schichtaufbau nach dem Kontaktlochätzen bei VFTL	9
1.10	Fototechnik und Ätzung der Gräben	9
1.11	Damascene-Struktur nach Metallisierung und Planarisierung	10
1.12	Klassischer Plattenkondensator und parasitäre Kapazität	10
1.13	k-Wert in Abhängigkeit von der Porosität	12
1.14	Ferminiveau in Metallen	14
1.15	Ferminiveau in dotierten Halbleitern	14
1.16	Bändermodell vor dem Kontakt von Metall und n-Silicium	15
1.17	Bändermodell nach dem Kontakt von Metall und n-Silicium	15
1.18	Bändermodell nach n^+ -Dotierung	16
1.19	Bändermodell nach dem Kontakt von Metall und p-Silicium	17
1.20	Bändermodell am Übergang p- und n-dotierter Siliciumkristalle	18
1.21	Prozessschritte beim Reflowrückätzen	19
1.22	Schematische Darstellung einer CMP-Anlage	20
1.23	CMP-Prozess zur Grabenisolation	21

Tabellenverzeichnis

1.1 Übersicht verschiedener organischer Siliciumoxide 13

1 Metallisierung

1.1 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 auszuhelfen.

1.2 Aluminiumtechnologie

1.2.1 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.

1.2.2 Siliciumdiffusion

Bei der Verwendung von reinem Aluminium kann es zu einer Diffusion von Siliciumatomen in das Metall kommen. Der Halbleiter reagiert bereits bei $200 - 250^\circ\text{C}$ 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.

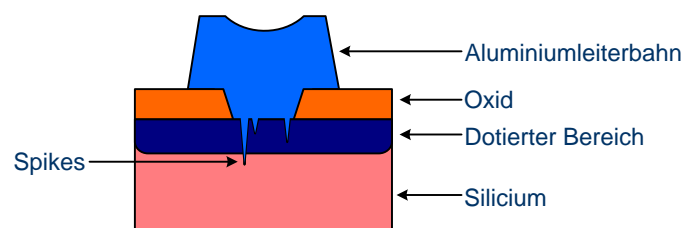


Abb. 1.1: Spikebildung

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.

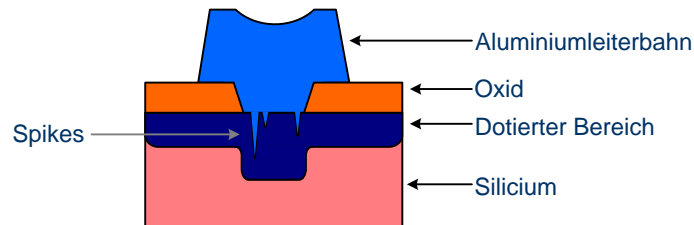


Abb. 1.2: Kontaktimplantation

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.

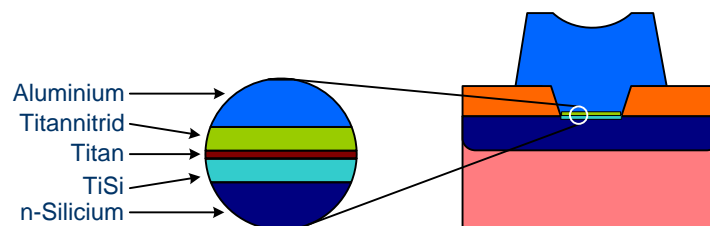


Abb. 1.3: Barrierschicht zwischen Aluminium und Substrat

1.2.3 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.

1.2.4 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, Titanitrid).

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:

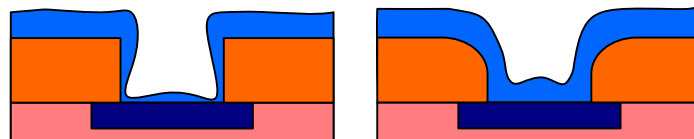


Abb. 1.4: Kantenverrundung zur konformen Aluminiumabscheidung

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.

1.3 Kupfertechnologie

1.3.1 Kupfertechnologie

Ab einer Strukturgröße von weniger als $250nm$ erfüllt Aluminium auch mit Kupferanteilen kaum noch die benötigten Anforderungen zur Verwendung in integrierten Schaltkreisen. So ist der elektrische Widerstand von Kupfer wesentlich geringer als in Aluminium, bei dem gut 50 % der zugeführten Leistung in Wärme umgewandelt werden. Auch die Stress- und Elektromigration sind in Aluminium wesentlich stärker ausgeprägt. Ein Umstieg auf Kupfer ist bei der stetigen Miniaturisierung somit unausweichlich.

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, verläuft im Wesentlichen in folgenden Schrit-

ten:

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

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

1.3.2 Damascene-Verfahren

Beim Damascene-Verfahren werden in bereits bestehende Zwischenschichten, die zur Isolierung bzw. Passivierung dienen, die Kontaktlöcher (VIA = Vertical Interconnect Access) der einzelnen Metallisierungsebenen geätzt, sowie die Gräben (Trenches), in denen später die Kupferleiterbahnen verlaufen sollen, strukturiert. In die Öffnungen kann dann mittels elektrochemischer Verfahren (Galvanotechnik) Kupfer abgeschieden werden. Ebenso sind auch CVD- oder PVD-Abscheidungen mit Reflowtechniken möglich. Anschließend wird das Kupfer in einem CMP-Prozess planarisiert, bis die Oberfläche eingeebnet ist.

Man unterscheidet dabei 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.

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 zum 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.

1. Der Wafer wird mit Resist beschichtet, der in einem Lithografieprozess strukturiert wird.

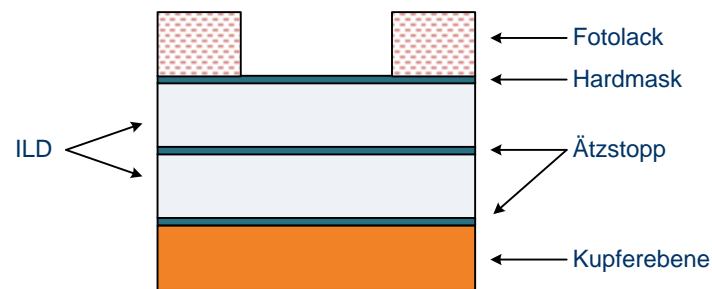


Abb. 1.5: Prinzipieller Damascene-Schichtaufbau vor der Strukturierung

2. In einem anisotropen Ätzprozess wird durch die Hartmaske (SiN) und die ILD-Schicht bis zur ersten Ätzstoppschicht (ebenfalls SiN) geöffnet.

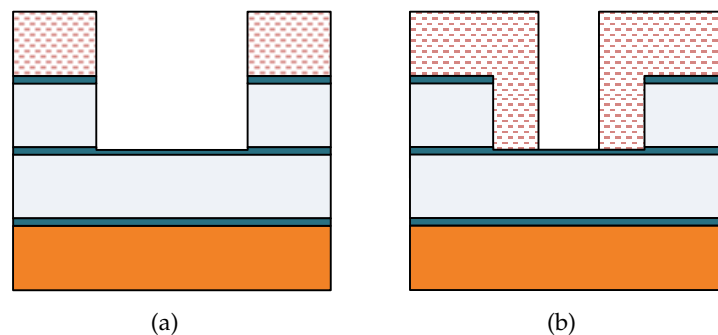


Abb. 1.6: Grabenätzung und VIA-Lackmaske

Der Fotoresist wird entfernt, und der Graben für die Leiterbahnen ist fertig strukturiert.

Die Hartmaske an der Oberfläche wird benötigt, um die ILD-Schicht während des Lackentfernens vor dem Plasma zu schützen. Dies ist notwendig, da das ILD 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.

4. Anschließend werden in einem anisotropen Ätzprozess die Kontaktlöcher (VIA) strukturiert.

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 und eine dünne Barriere-

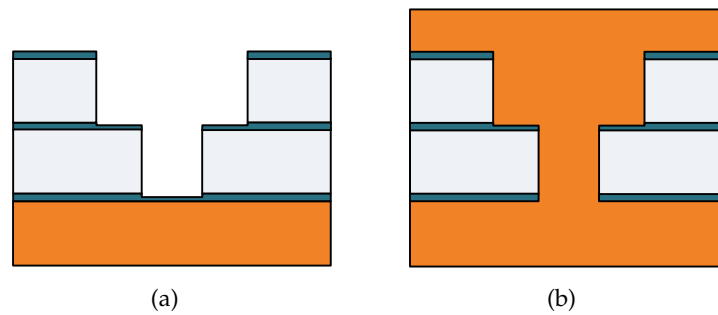


Abb. 1.7: Finale Struktur nach dem Ätzen und der Kupferabscheidung

schicht aus Tantal abgeschieden die verhindert, dass das anschließend aufgebraachte Kupfer in das ILD eindringt.

5. Eine dünne Kupferkeimschicht wird aufgebracht und die Strukturen in einem galvanischen Verfahren mit Kupfer aufgefüllt.
6. Das Kupfer wird abschließend durch chemisch mechanisches Polieren planarisiert.

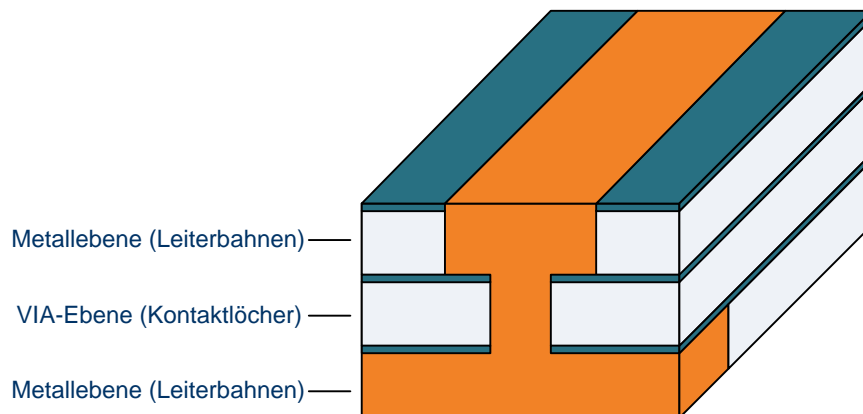


Abb. 1.8: Damascene-Struktur nach Metallisierung und Planarisierung

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

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, damit das darunter befindliche Kupfer nicht herausgeschlagen wird und in das ILD eindiffundiert.

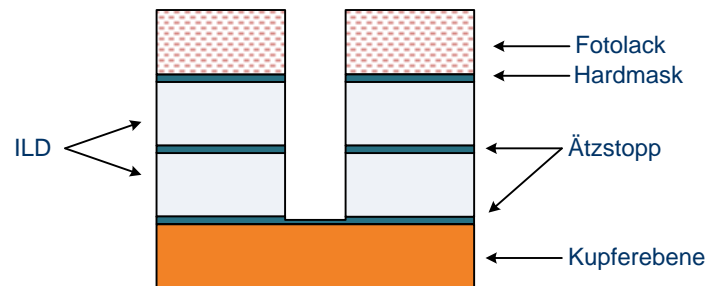


Abb. 1.9: Damascene-Schichtaufbau nach dem Kontaktlochätzen bei VFTL

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

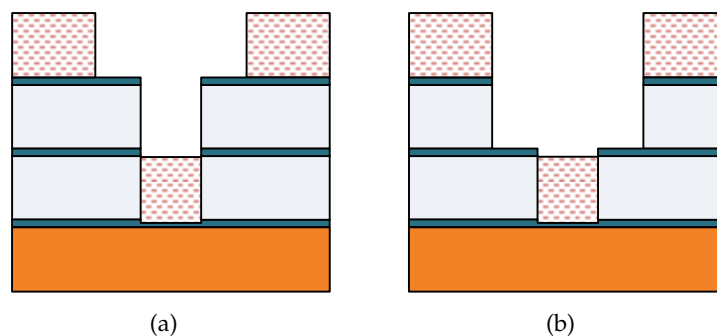


Abb. 1.10: Fototechnik und Ätzung der Gräben

3. Der unterste Ätzstopp wird durch den aufgebrachtten 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.

4. Als letztes folgt die Kupferabscheidung und die Planarisierung mittels CMP.

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

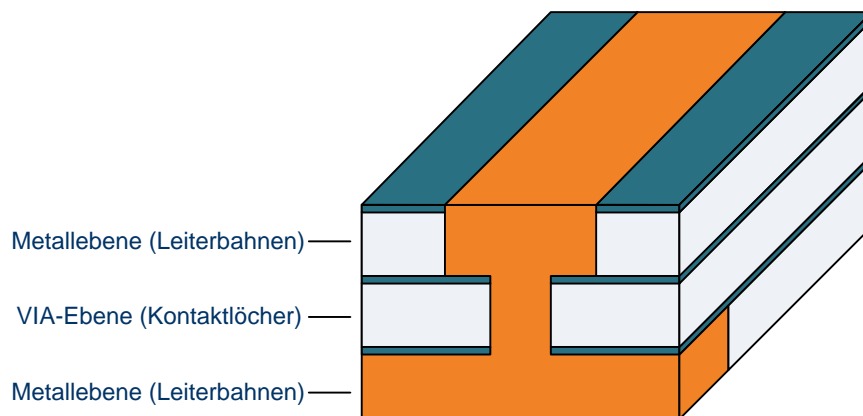


Abb. 1.11: Damascene-Struktur nach Metallisierung und Planarisierung

1.3.3 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.

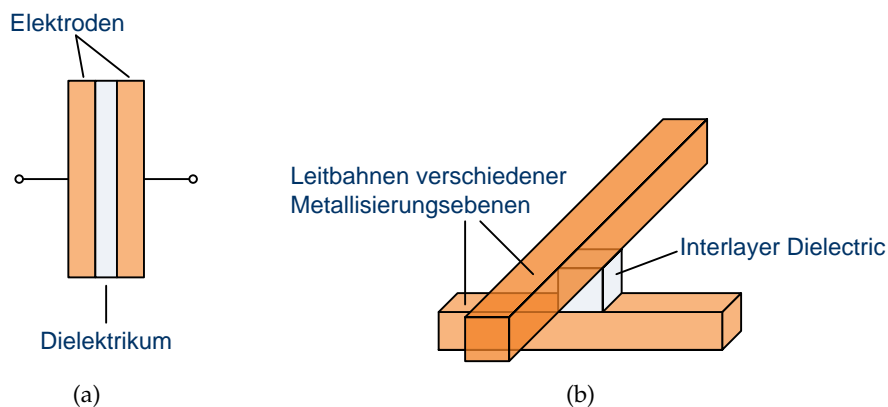


Abb. 1.12: (a) Klassischer Plattenkondensator und (b) Kapazität zwischen Leiterbahnen

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 Verringerung 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 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 Polarisation (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 verringern, gibt es im Grunde zwei Ansätze:

- die Polarisierbarkeit von Bindungen im Dielektrikum verringern
- die Anzahl an Bindungen 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. Dadurch verringert sich ϵ_r für die gesamte Schicht. Die Poren

können erzeugt werden, indem das ILD-Material mit Polymeren versetzt wird, die in einem Temperschritt aus der Schicht getrieben werden.

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

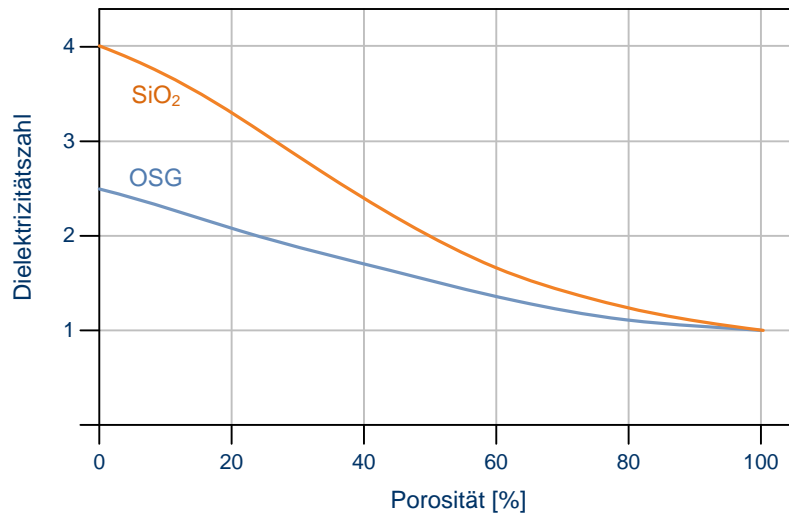


Abb. 1.13: k-Wert in Abhängigkeit von der Porosität

Ebenso können Prozessgase oder Kupfer aus den Leiterbahnen leichter in die poröse 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 (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. 1.1: Übersicht verschiedener organischer Siliciumoxide

1.4 Der Metall-Halbleiter-Kontakt

1.4.1 Kontaktierung von dotierten Halbleitern

Nach der Herstellung der Transistoren im Siliciumsubstrat müssen diese mittels elektrischer Kontakte miteinander verbunden werden. Dabei wird zum einen die Gateelektrode zur Steuerung des Transistors kontaktiert, zum anderen müssen die dotierten Source- und Draingebiete, über die der Stromfluss erfolgt, angesteuert 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 Fermi-niveau eine wichtige Rolle. Das Fermi-niveau ist das Energieniveau, bis zu dem sich am absoluten 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 kann die Wasseroberfläche eines Sees betrachtet werden. Die Wassermoleküle darunter stellen die Elektronen dar, welche bis an die Oberfläche – das Fermi-niveau – reichen.

In dotierten Halbleitern befinden sich Fremdatome als Donatoren oder Akzeptoren im Kristallgitter. In n-dotierten Halbleitern befindet sich das Fermi-niveau in der Nähe der Leitungsbandkante, da die Donatoratome schon bei geringer Energiezufuhr freie Elektronen zur Verfügung stellen können. Dementsprechend befindet sich das Fermi-niveau in einem p-dotierten Halbleiter in der Nähe der Valenzbandkante, da Elektronen aus dem Valenzband des Siliciumkristalls leicht vom Akzeptoratom aufgenommen werden können (vgl. Kapitel Dotieren).

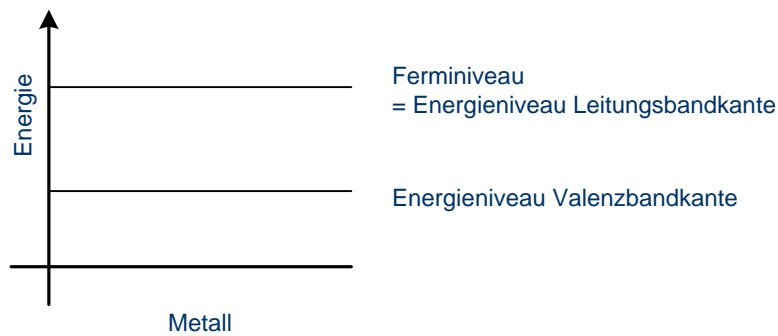


Abb. 1.14: Ferminiveau in Metallen

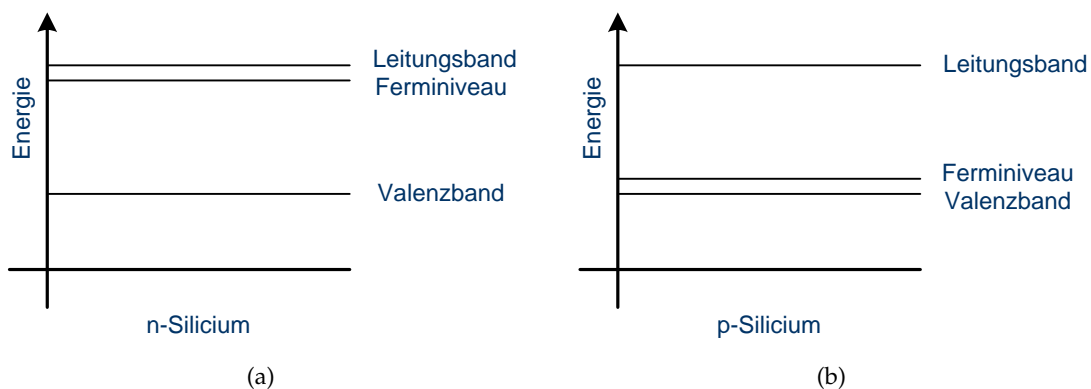


Abb. 1.15: Ferminiveau in dotierten Halbleitern

1.4.2 n-Halbleiter-Kontakt

Da das Leitungsband in n-Halbleitern energetisch höher liegt als das Ferminiveau, können bei der Kontaktierung Elektronen aus dem Silicium in das Metall fließen, da diese den energetisch niedrigsten Zustand einnehmen.

Es verringert sich somit die Aufenthaltswahrscheinlichkeit von Elektronen im Leitungsband des Halbleiters, und das Ferminiveau, 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) welche die verbliebenen Elektronen im n-Leiter überwinden müssen, um in das Metall zu fließen.

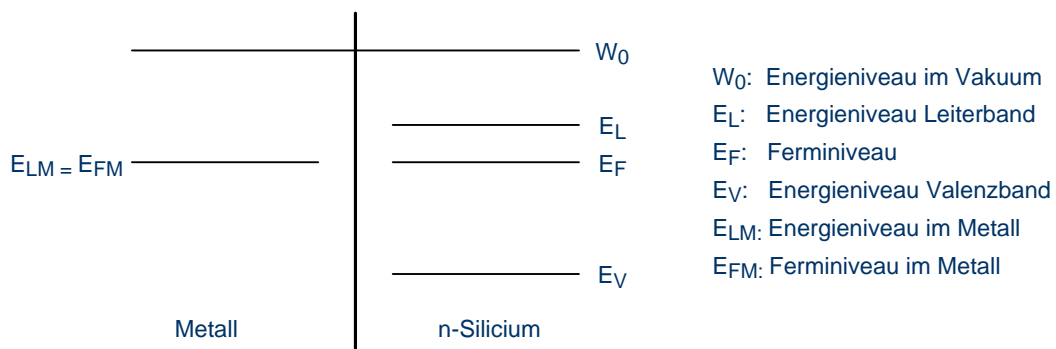


Abb. 1.16: Bändermodell vor dem Kontakt von Metall und n-Silicium

Beim Kontakt von Metall und Halbleiter gleichen sich die Fermienergien also durch Diffusionsprozesse an, im Bereich der Grenzfläche ist das Fermienergie konstant.

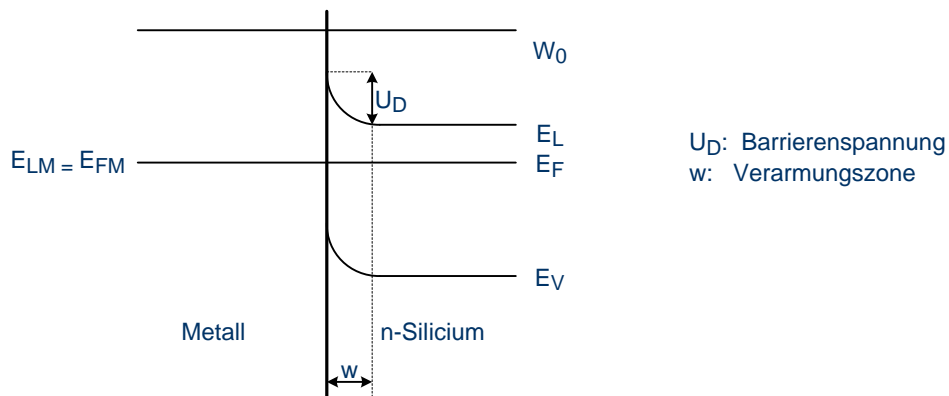


Abb. 1.17: Bändermodell nach dem Kontakt von Metall und n-Silicium

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.

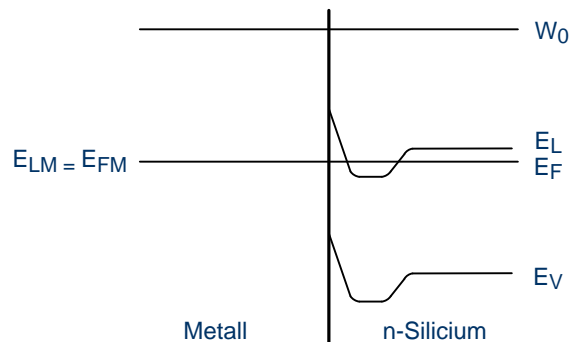


Abb. 1.18: Bändermodell nach n^+ -Dotierung

Da Aluminium im Silizium als Elektronenakzeptor eingebaut wird (es nimmt Elektronen auf) und sich so eine p-Dotierung an der Grenzfläche bildet, entsteht bei p-dotiertem Silizium 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 Silizium) an der Kontaktfläche aufgebracht werden.

Im Gegensatz zur Diode beim p-n-Übergang, bei der die Schaltgeschwindigkeit auf der Diffusion von Elektronen beruht, haben Schottky-Dioden sehr kurze Schaltzeiten. Sie eignen sich daher als Schutzdioden um Spannungsspitzen abzufangen.

1.4.3 p-Halbleiter-Kontakt

Bei Metall-p-Halbleiter-Kontakten ergibt sich in Folge des Ladungsträgeraustauschs zwischen Metall und Halbleiter eine Bandverbiegung nach unten, Löcher im Halbleiter rekombinieren mit Elektronen aus dem Metall. Durch die Verringerung der Löcherkonzentration ergibt sich eine negative Raumladungszone im Halbleiterkristall, das Fermi-niveau – repräsentativ für die höchsten Besetzungszustände durch Elektronen – entfernt sich und die so entstandene Potentialbarriere an der Valenzbandkante verhindert eine weitere Bewegung der Löcher, welche – komplementär zu Elektronen – die energetisch höchsten Zustände einnehmen wollen.

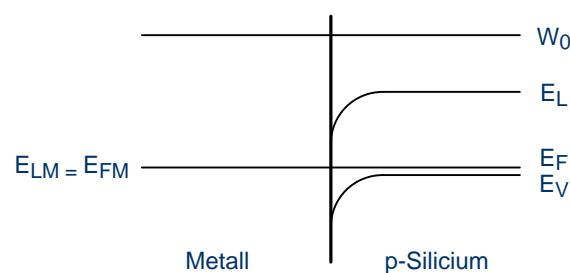


Abb. 1.19: Bändermodell nach dem Kontakt von Metall und p-Silicium

Ohne eine äußere Spannung kommen die Diffusionsprozesse zum Erliegen. Auch hier gleichen sich die Fermi-niveaus im thermodynamischen Gleichgewicht einander an.

1.4.4 Bändermodell eines p-n-Übergangs

Aus der Tatsache, dass das Fermi-niveau konstant sein muss (andernfalls würden Elektronen an Orte mit niedrigerem Fermi-niveau fließen, dort freie Zustände besetzen und damit das Fermi-niveau wieder anheben), 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. $0,7V$.

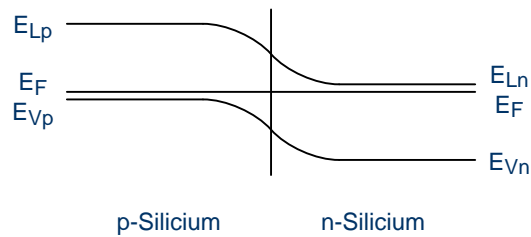


Abb. 1.20: Bändermodell am Übergang p- und n-dotierter Siliciumkristalle

1.5 Mehrlagenverdrahtung

1.5.1 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.

1.5.2 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.

1.5.3 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.

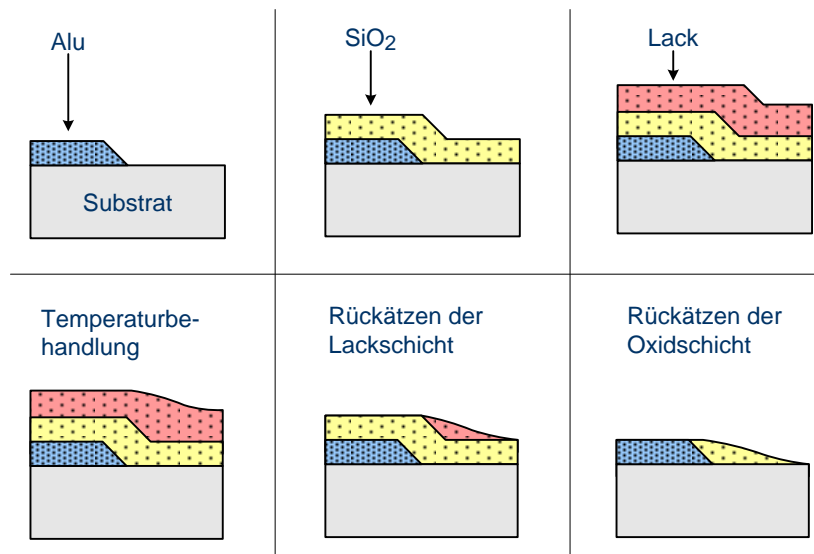


Abb. 1.21: Prozessschritte beim Reflowrückätzen

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.

1.5.4 Chemisch-Mechanisches Polieren

Beim chemisch mechanischen Polieren (auch chemisch mechanisches Planarisieren, kurz CMP) wird im Gegensatz zu den Reflowtechniken eine Homogenität über die ge-

samte 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 Vakuumanströmung (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 zu 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.

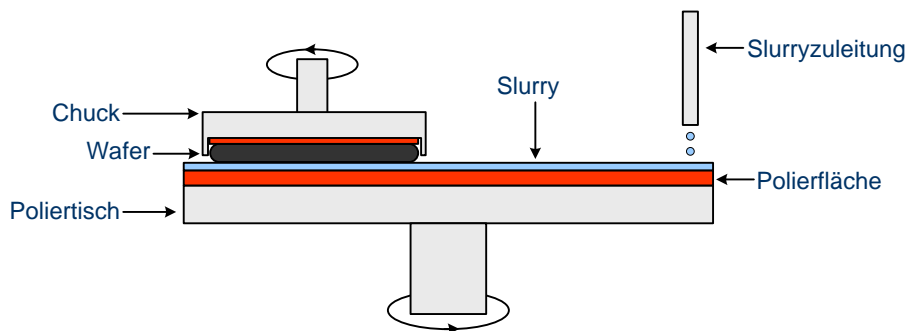


Abb. 1.22: Schematische Darstellung einer CMP-Anlage

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.

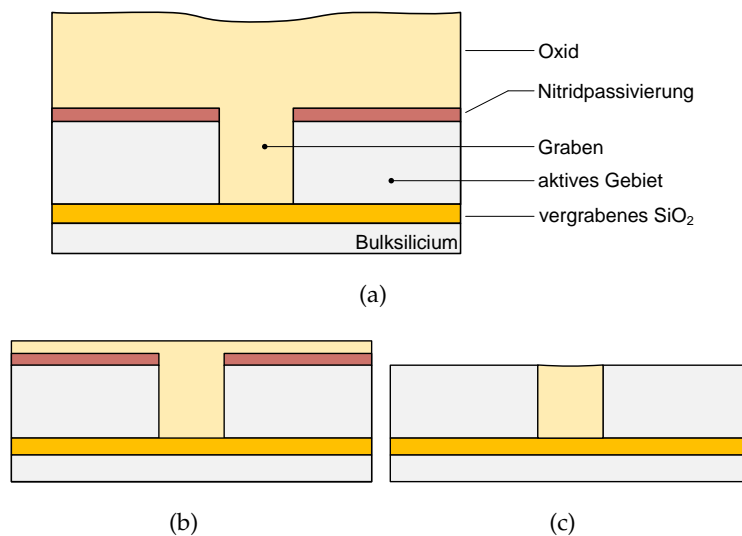


Abb. 1.23: CMP-Prozess zur Grabenisolation: (a) Vor dem Polieren, (b) nach dem ersten Polierschritt, (c) nach dem finalen Polierschritt und Nitridentfernen

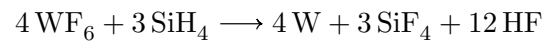
Auch wenn dieses Verfahren recht grob anmutet, lässt sich hierbei doch eine auf wenige Nanometer planare Oberfläche herstellen.

1.5.5 Kontaktierung der Metallisierungsebene

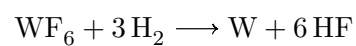
Um die Metallebenen zu verbinden, werden in die Isolationsschichten Kontaktlöcher mit sehr hoher Anisotropie geätzt, so werden Kanten an den Kontaktlöchern vermieden. Die Kontaktlöcher müssen so aufgefüllt werden, dass einerseits eine optimale Kontaktierung gewährleistet wird und zugleich die Oberfläche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

