

Halbleitertechnologie von A bis Z

Metallisierung

Mehrlagenverdrahtung	3
Mehrlagenverdrahtung	3
BPSG-Reflow	3
Reflowrückätzen	4
Chemisch mechanisches Polieren	5
Kontaktierung der Metallisierungsebenen	5

Mehrlagenverdrahtung

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. In einem heutigen Prozessor summieren sich die Leiterbahnen auf mehrere zehn Kilometer Gesamtlänge.

Zwischen den Verdrahtungsebenen sind Isolationsschichten aufgebracht, durch Kontaktöffnungen (VIA, vertical interconnect access) werden die einzelnen Ebenen miteinander verbunden. Gebräuchlich sind heute etwa zehn bis zwanzig Verdrahtungsebenen.

Diese Ebenen sind nicht gleich aufgebaut, sondern nach ihrer Aufgabe gestaffelt:

- Lokale Verdrahtung: Die untersten Ebenen verbinden benachbarte Transistoren über kurze Wege. Sie haben die feinsten Abmessungen, den größten Widerstand je Länge und werden mit den aufwendigsten Lithografieverfahren hergestellt.
- Mittlere Ebenen: Sie verbinden Funktionsblöcke innerhalb des Chips; Breite und Dicke der Leiterbahnen nehmen mit jeder Ebene zu.
- Globale Verdrahtung: Die obersten Ebenen führen Takt und Versorgungsspannung über den ganzen Chip. Sie sind um ein Vielfaches breiter und dicker als die untersten, weil sie große Ströme führen und dabei möglichst wenig Spannung abfallen darf.

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. Hinzu kommt, dass die **Belichtung** feiner Strukturen nur eine geringe Schärfentiefe hat und deshalb eine ebene Oberfläche voraussetzt. Um die Kanten bzw. Stufen zu entfernen gibt es mehrere Möglichkeiten zur Planarisierung, die im folgenden erläutert werden.

BPSG-Reflow

Bei der Reflowtechnik werden Schichten aus dotierten Gläsern auf dem Wafer aufgebracht. Weit verbreitet sind Phosphorsilicatglas (PSG) und

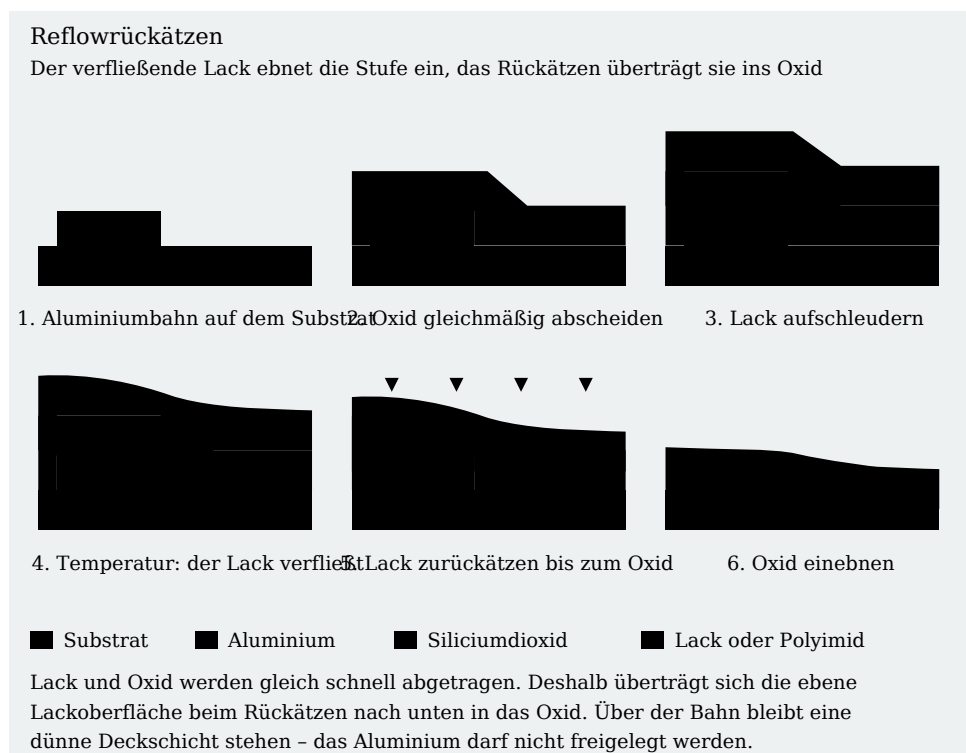
Borophosphorsilicatglas (BPSG). In einem Hochtemperaturschritt verfließen die Gläser und bilden eine ebene Oberfläche: bei PSG und BPSG geschieht dies bei ca. 900 °C. Zur Planarisierung auf einer Verdrahtungsebene ist diese Technik jedoch nicht geeignet, da das Aluminium unter den hohen Temperaturen aufschmelzen würde.

Damit ist der Einsatzbereich eng: Das Verfahren kommt nur vor der ersten Metallebene in Frage, solange noch kein Metall auf der Scheibe liegt. Auch dort ist es weitgehend abgelöst worden, weil die Temperatur die bereits eingebrachten Dotierprofile verändert und weil das Verfließen nur örtlich einebnet – über die ganze Scheibe hinweg bleibt die Oberfläche uneben. Beides leistet das [chemisch mechanische Polieren](#) besser.

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.



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.

Genau daran sind sie gescheitert. Sie glätten die Umgebung einer einzelnen Stufe, lassen aber die großräumigen Höhenunterschiede über die Scheibe hinweg bestehen – und gerade die stören die Belichtung feiner Strukturen. Beide Verfahren sind deshalb durch das chemisch mechanische Polieren abgelöst worden und heute nur noch von historischem Interesse.

Chemisch mechanisches Polieren

Nach dem Damascene-Verfahren wird die überschüssige Kupferschicht und die Barriere per chemisch-mechanischem Polieren (CMP) bis auf das Zwischenoxid abgetragen, sodass isolierte Leiterbahnen zurückbleiben. Da dieser Schritt auch bei jeder anderen Ebene des Wafers zum Einsatz kommt, wird er im eigenen Kapitel ausführlich behandelt.

[Mehr im Kapitel Chemisch-Mechanisches Polieren \(CMP\) →](#)

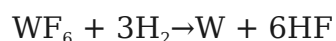
Kontaktierung der Metallisierungsebenen

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:



Der Vorteil des Verfahrens liegt darin, dass die Abscheidung aus der Gasphase auch enge, tiefe Löcher gleichmäßig von innen heraus füllt, während ein

aufgedampftes oder gesputtertes Metall die Öffnung oben zuwachsen ließe und einen Hohlraum einschliesse. Wolfram wird deshalb ganzflächig abgeschieden und anschließend per CMP so weit zurückpoliert, dass es nur noch in den Löchern steht.

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.

Der Kontakt als Engstelle

Mit kleiner werdenden Abmessungen verschiebt sich das Problem. Der Widerstand eines Kontakts setzt sich aus dem Widerstand des Füllmetalls, dem der Barriere und dem Übergangswiderstand zum Silicium zusammen. Die ersten beiden Anteile wachsen mit schrumpfendem Querschnitt schnell an, weil Wolfram eine vergleichsweise dicke Haftschrift aus Titan und Titannitrid benötigt, die selbst kaum leitet. In den untersten Ebenen moderner Prozesse wird Wolfram deshalb durch Cobalt oder Ruthenium ersetzt: Beide haben zwar einen höheren spezifischen Widerstand als Wolfram, kommen aber mit einer sehr dünnen oder ganz ohne Haftschrift aus. Bei den kleinsten Kontakten leitet die Füllung dadurch insgesamt besser, obwohl das Material für sich genommen schlechter leitet.