Halbleitertechnologie von A bis Z

Trockenätzen

Übersicht 3
Allgemein 3
Wichtige Größen beim Trockenätzen 3
Anregung und Verfahrensarten 4
Trockenätzverfahren 5
Ionenstrahlätzen 5
Plasmaätzen 7
Reaktives Ionenätzen 10
Eingesetzte Ätzgase beim Trockenätzen 14
Sonderfall: Bosch-Prozess (Deep Reactive Ion Etching) 16
Anlagentechnik (Reaktortypen) 16
Allgemein: Aufbau einer Trockenätzanlage 16
Reaktortypen: CCP und ICP in der Praxis 16
Elektronenzyklotronresonanz (ECR) 17
Wafer-Chuck: Temperaturkontrolle und elektrostatische Klemmung 18
Ätzen von Verbindungshalbleitern und High-k-Materialien 18
Warum diese Materialien eine Sonderrolle spielen 18
III-V-Verbindungshalbleiter (GaAs, InP, GaN) 19
Siliciumcarbid (SiC) 19
High-k-Dielektrika (HfO_2 & Co.) 19

Übersicht

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 der meisten 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.

Wichtige Größen beim Trockenätzen

Trockenätzprozesse werden über einige wenige Kenngrößen beschrieben, die zugleich die Zielgrößen jeder Prozessentwicklung sind.

Ätzrate r: Die Ätzrate beschreibt den Ätzabtrag pro Zeit und wird z.B. in Nanometer pro Minute oder Ångström pro Sekunde angegeben.

$$r=\frac{Ätzabtrag \Delta z}{Ätzzeit \Delta t}$$

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}: Die Selektivität S_{jk} zwischen Material j und Material k beschreibt das Verhältnis der Ätzraten 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 Reflowrückätzen muss die Selektivität 1 betragen um Topografien gleichmäßig einzuebnen.

Uniformität U: Die Uniformität beschreibt, wie gleichmäßig der Abtrag ausfällt –

über eine Scheibe hinweg, von Scheibe zu Scheibe und von Charge zu Charge. Sie wird als Streuung der Ätzrate bezogen auf deren Mittelwert angegeben.

$$U=\frac{r_{max}-r_{min}}{r_{max}+r_{min}}\cdot 100\,\%$$

Aspektverhältnis A: Das Aspektverhältnis ist das Verhältnis von Ätztiefe z zu Strukturbreite b. Es beschreibt, wie schlank ein Graben oder Loch ist.

$$A=\frac{z}{b}$$

Mit steigendem Aspektverhältnis sinkt die Ätzrate, weil die Ätzteilchen schlechter an den Strukturboden gelangen und die Reaktionsprodukte schlechter abgeführt werden. Dieser Effekt wird als aspektverhältnisabhängiges Ätzen (aspect ratio dependent etching, ARDE) oder RIE-Lag bezeichnet. Er ist heute die maßgebliche Grenze beim Ätzen tiefer Strukturen, etwa der Speicherlöcher im 3D-NAND-Flash mit Aspektverhältnissen von mehr als 50:1. Typische Profilfehler solcher Ätzungen sind das Bowing (bauchige Aufweitung im oberen Bereich), das Twisting (Verkippen tiefer Löcher gegeneinander) und das Microtrenching (überhöhter Abtrag in den Bodenecken).

Loading-Effekt: Die Ätzrate hängt von der insgesamt freiliegenden Fläche ab: Eine große zu ätzende Fläche verbraucht mehr Reaktivteilchen, die Ätzrate sinkt. Makroloading betrifft die gesamte Scheibe bzw. Charge, Mikroloading die unmittelbare Umgebung einer einzelnen Struktur.

Da die Ätzrate von Anlagenzustand, Gasfluss und Belegung abhängt, wird eine Ätzung in der Fertigung nicht über eine feste Zeit gesteuert, sondern über eine Endpunkterkennung. Dazu verfolgt ein Spektrometer die optische Emission des Plasmas (optical emission spectroscopy, OES) – beim Durchätzen einer Schicht ändert sich die Intensität charakteristischer Linien, weil andere Reaktionsprodukte entstehen – oder ein Interferometer misst die verbleibende Schichtdicke. Nach dem erkannten Endpunkt folgt eine kurze, definierte Überätzzeit, um Reste über die gesamte Scheibe hinweg sicher zu entfernen.

Anregung und Verfahrensarten

Bei den Trockenätzverfahren werden Gase durch hochfrequente Wechselfelder angeregt, typisch sind 13,56 MHz und 2,45 GHz. Bei einem Druck zwischen 0,1 Pa und 100 Pa liegt die freie Weglänge der Teilchen bei wenigen Millimetern bis Zentimetern.

Moderne Anlagen trennen die Erzeugung des Plasmas von der Beschleunigung der Ionen: Ein Generator hoher Frequenz (z.B. 13,56 MHz oder 60 MHz, induktiv oder per Mikrowelle eingekoppelt) stellt die Ladungsträgerdichte und damit die Ätzrate ein, ein zweiter Generator niedriger Frequenz (z.B. 400 kHz bis 2 MHz) an der Waferelektrode die Ionenenergie und damit den physikalischen Anteil der

Ätzung. Beide Größen sind so weitgehend unabhängig voneinander einstellbar. Zusätzlich wird das Plasma häufig gepulst, um Ladungsansammlungen und Profilfehler in tiefen Strukturen zu verringern.

Es gibt generell drei verschiedene Arten des Trockenätzens, welche auf der folgenden Seite näher beschrieben werden:

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

Als jüngere Entwicklung kommt das atomlagengenaue Ätzen (atomic layer etching, ALE) hinzu. Dabei werden die chemische Aktivierung der Oberfläche und der eigentliche Abtrag zeitlich getrennt und zyklisch wiederholt, so dass pro Zyklus nur eine oder wenige Atomlagen entfernt werden. Der Abtrag wird damit nicht über die Ätzzeit, sondern über die Anzahl der Zyklen eingestellt – eine Voraussetzung für Bauelemente, deren Schichtdicken nur noch wenige Nanometer betragen, etwa Gate-all-around-Transistoren.

Trockenätzverfahren

Ionenstrahlätzen

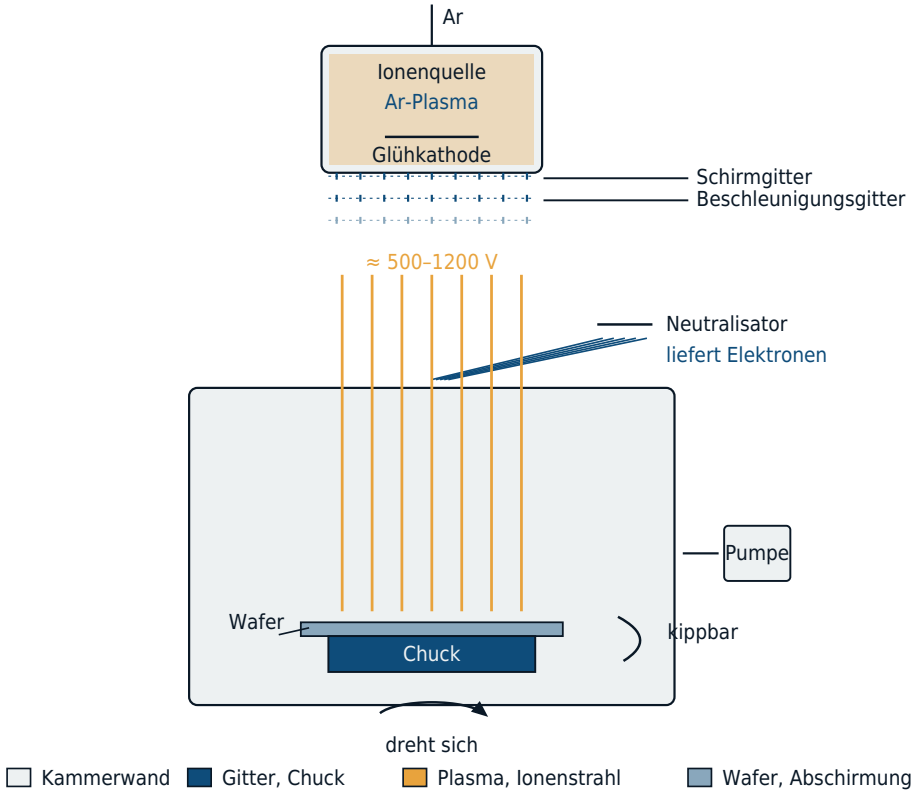
Das Ionenstrahlätzen (ion beam etching, IBE) ist ein rein physikalisches Trockenätzverfahren. Dabei werden Argonionen als gerichteter Ionenstrahl mit 1–3 keV auf die Scheibe gelenkt. Durch die Energie der Teilchen schlagen diese Material aus der Oberfläche heraus. Die Ionen werden in einer vom Prozessraum getrennten Quelle erzeugt und über ein Gittersystem extrahiert; der Arbeitsdruck liegt unterhalb von 0,1 Pa, damit die Ionen ihre Richtung bis zur Scheibe behalten. Der Wafer wird 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.

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. Wird das reaktive Gas bereits in der Ionenquelle zugegeben, spricht man von reaktivem Ionenstrahlätzen (reactive ion beam etching, RIBE), wird es erst vor der Scheibe eingeleitet, von chemisch unterstütztem Ionenstrahlätzen (chemically assisted ion beam etching, CAIBE).

Ionenstrahl-Ätzreaktor

Ein kollimierter Ionenstrahl trägt das Material rein physikalisch ab



Das Gittersystem beschleunigt die Argon-Ionen zu einem gebündelten Strahl; der Neutralisator liefert Elektronen, damit sich der Wafer nicht auflädt. Weil kein reaktives Gas beteiligt ist, wird nahezu jedes Material rein mechanisch abgetragen – dafür fehlt die chemische Selektivität des reaktiven Ionenätzens.

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 Ionendichte im Strahl) spielt das Ionenstrahlätzen bei der Strukturierung von Silicium und Siliciumverbindungen keine Rolle mehr. Unverzichtbar ist es dagegen bei Materialien, die keine flüchtigen Reaktionsprodukte bilden und daher chemisch kaum zu ätzen sind:

- magnetische Schichtstapel, etwa die Tunnelelemente von MRAM-

Speicherzellen aus CoFeB und MgO, sowie Leseköpfe und Magnetfeldsensoren
- Edelmetalle wie Platin, Iridium und Ruthenium
- piezoelektrische und ferroelektrische Schichten wie PZT
- Lithiumniobat und ähnliche Kristalle in der integrierten Photonik

Der kippbare und rotierende Substrathalter ist dabei ein wesentlicher Vorteil: Über den Einfallswinkel lassen sich der Kantenwinkel der Struktur einstellen und wieder angelagertes Material von den Seitenwänden entfernen – bei Schichtstapeln aus vielen unterschiedlichen Materialien ein Ergebnis, das mit plasmachemischen Verfahren nicht erreichbar ist. Da isolierende Substrate sich unter Ionenbeschuss aufladen würden, arbeitet die Quelle mit einem Neutralisator, der dem Strahl Elektronen zusetzt.

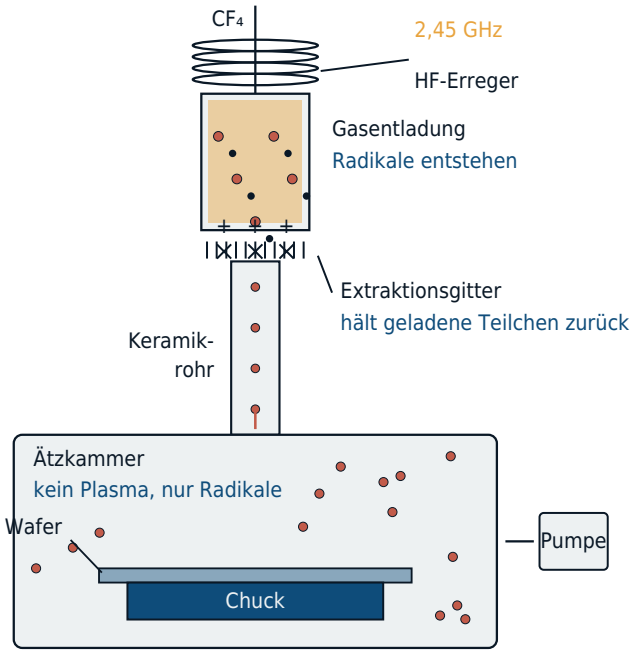
Plasmaätzen

Das Plasmaätzen ist ein rein chemisches Ätzen (engl. chemical dry etching, CDE). 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.

Eine Anlagenart beim Plasmaätzen ist der Down-Stream-Reaktor. Dabei wird durch Anlegen einer hochfrequenten Spannung (z.B. 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.

CDE-Downstream-Reaktor

Die Gasentladung brennt getrennt vom Wafer – nur Radikale erreichen ihn



Gasentladung Radikale + geladenes Teilchen Chuck Wafer

Weil das Plasma vom Wafer getrennt brennt, treffen keine beschleunigten Ionen auf die Scheibe – nur ungeladene, aber reaktionsfreudige Radikale. Das Ätzen bleibt dadurch schädigungsfrei, aber isotrop: Es greift in alle Richtungen an.

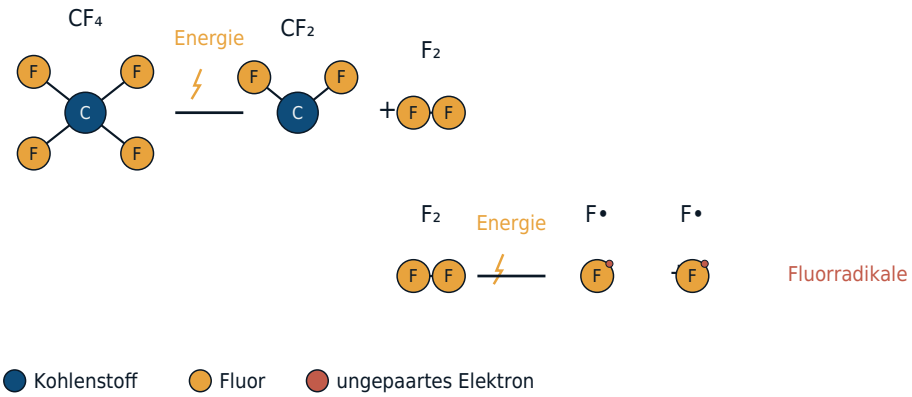
In der Zone der Gasentladung entstehen durch die Stöße der Gasmoleküle untereinander verschiedene Teilchen, unter anderem Radikale. Radikale sind Moleküle mit aufgespaltener Bindung, die durch ein ungepaartes Außenelektron sehr reaktionsfreudig sind. Als neutrales Gas wird beispielsweise Tetrafluormethan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

$$2CF_4 + O_2 \rightarrow 2COF_2 + 2F_2$$

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

Bildung von Fluorradikalen

Energie in der Gasentladung spaltet CF_4 in zwei Schritten auf



Ein Radikal ist ein ungeladenes Teilchen mit einem ungepaarten Außenelektron – das macht es sehr reaktionsfreudig. Zusätzlich lässt sich Fluor auch mit zugesetztem Sauerstoff abspalten: $2\ CF_4 + O_2 \rightarrow 2\ COF_2 + 2\ F_2$.

Neben den neutralen Radikalen entstehen weitere, teils geladene Teilchen ($CF4^+$, $CF3^+$, $CF2^+$, ...), welche gemeinsam über das Gasrohr zur Ätzkammer geleitet werden. 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, wo sie mit der Scheibenoberfläche reagieren. 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: $Si + 4F \rightarrow SiF_4$
Siliciumdioxid: $SiO_2 + 4F \rightarrow SiF_4 + O_2$
Siliciumnitrid: $Si_3N_4 + 12F \rightarrow 3SiF_4 + 2N_2$

Heutige Bedeutung

Zur Strukturierung von Schichten ist das rein chemische Trockenätzen ungeeignet, weil das isotrope Profil die Maske unterätzt. Als Remote- oder Downstream-Prozess – das Plasma brennt getrennt von der Scheibe, nur langlebige neutrale Radikale erreichen sie – erfüllt es in der Fertigung aber klar umgrenzte Aufgaben:

- Lackentfernung (Ashing): Sauerstoffradikale verbrennen den Fotolack nach der Ätzung oder Implantation zu CO_2 und H_2O.
- Kammerreinigung: Stickstofftrifluorid NF_3 wird in einer vorgeschalteten Plasmaquelle zerlegt; die Fluorradikale entfernen Abscheidungen aus Prozesskammern, ohne dass diese geöffnet werden müssen.

- Selektives Rückätzen: Radikalprozesse aus NH_3 und NF_3 entfernen natives Oxid und dünne Nitridschichten schädigungsfrei – auch in Kontaktlöchern, in die eine gerichtete Ätzung nur schwer hineinreicht.
- Freilegen von Kanälen: Beim Nanosheet- bzw. Gate-all-around-Transistor wird Silicium-Germanium selektiv gegen Silicium isotrop entfernt, um die Kanalschichten freizustellen. Hier wird die Isotropie, die bei der Strukturierung stört, gezielt ausgenutzt.

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.

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 Radikalen 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

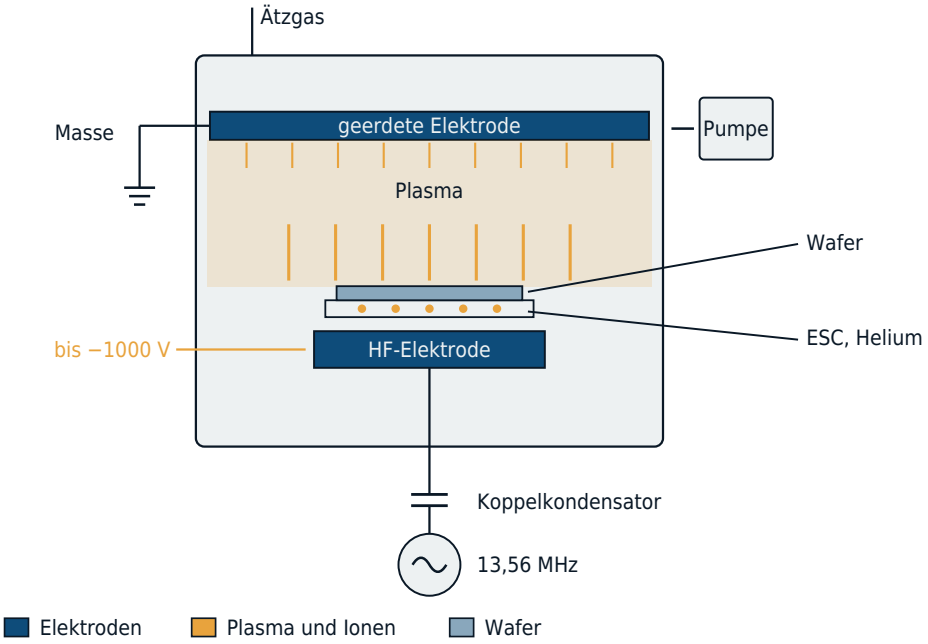
ändern. So geschieht keine gerichtete Ätzung mehr auf der Scheibenoberfläche, der Ätzprozess nimmt einen stärker chemischen Charakter an, das Ätzprofil ist isotrop, die Selektivität erhöht sich.

Die Hexodenbauform ätzte mehrere Scheiben gleichzeitig auf einem sechsseitigen Träger. Solche Chargenanlagen sind heute nur noch von historischem Interesse, weil sich Homogenität und Endpunkt über viele Scheiben hinweg nicht ausreichend genau kontrollieren lassen.

Heute wird ausschließlich einzelscheibenweise in Parallelplattenbauweise geätzt. Der Wafer liegt auf einem elektrostatischen Chuck (electrostatic chuck, ESC), der ihn plan festhält und über Helium auf der Scheibenrückseite thermisch an die temperierte Elektrode koppelt. Damit sind Wafertemperaturen von etwa −100 °C bis über 100 °C einstellbar – die Temperatur ist eine der wirksamsten Stellgrößen für Selektivität und Profil.

Reaktives Ionenätzen

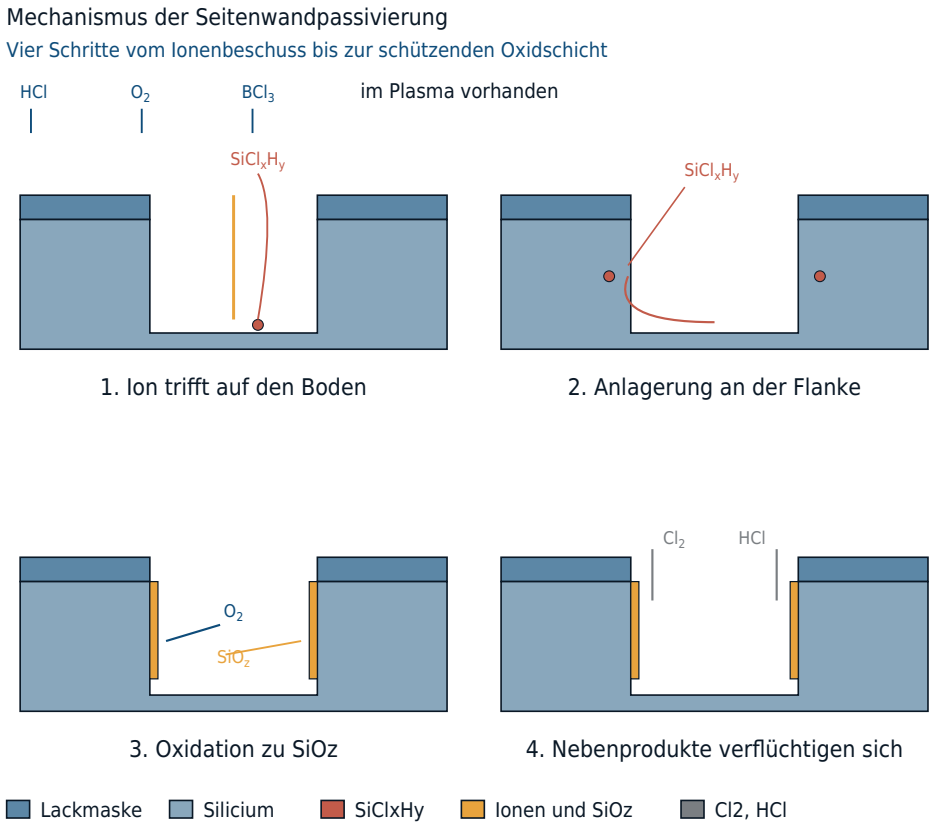
Parallelplattenbauweise: Der Wafer liegt auf der Hochfrequenzelektrode



Weil die untere Elektrode kleiner ist als die obere, lädt sie sich gegenüber dem Plasma negativ auf. Die positiven Ionen werden dadurch senkrecht beschleunigt und tragen gerichtet ab – das Ätzprofil wird anisotrop.

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. Elektrodentemperatur 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 wurde früher Kohlenstoff eingesetzt: 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. Heute sind Elektroden aus Silicium, Siliciumcarbid oder Quarz üblich; die dem Plasma zugewandten Kammerteile werden mit Yttrium- oder Aluminiumoxid beschichtet, um Partikel und metallische Verunreinigungen zu

vermeiden.

Weiterentwicklungen des reaktiven Ionenätzens

Der klassische RIE-Reaktor koppelt Plasmadichte und Ionenenergie über dieselbe Elektrode: Mehr Leistung bedeutet gleichzeitig mehr Ätzrate und mehr Schädigung. Die heutigen Anlagengenerationen lösen diese Kopplung auf und erweitern das Verfahren in mehrere Richtungen.

- Hochdichte Plasmaquellen: Bei induktiv gekoppelten Anlagen (inductively coupled plasma, ICP, auch TCP) erzeugt eine Spule über einem dielektrischen Fenster das Plasma, bei ECR-Anlagen eine Mikrowelle im Magnetfeld. Die Ionenenergie wird getrennt über einen Bias-Generator an der Waferelektrode eingestellt. Ergebnis sind hohe Ätzraten bei geringer Ionenenergie, also weniger Schädigung, bei Drücken von deutlich unter 1 Pa.
- Mehrfrequenz-Anlagen: Beim Ätzen von Dielektrika arbeiten kapazitiv gekoppelte Reaktoren mit zwei oder drei Generatoren unterschiedlicher Frequenz auf derselben Elektrode, um Ionendichte, Ionenenergie und Energieverteilung getrennt einzustellen. Für tiefe Strukturen werden Bias-Spannungen von mehreren Kilovolt eingesetzt.
- Gepulstes Plasma: Wird die eingespeiste Leistung im Kilohertzbereich getaktet, klingt der Ladungsaufbau in tiefen Strukturen zwischen den Pulsen ab. Das verringert Profilfehler wie Bowing und Notching und verbessert die Selektivität.
- Tiefenätzen (DRIE, Bosch-Prozess): Beim tiefen reaktiven Ionenätzen wechseln sich sehr kurze Ätzschritte mit SF_6 und Passivierschritte mit C_4F_8 zyklisch ab. So entstehen Gräben mit Aspektverhältnissen über 30:1 und Tiefen von einigen hundert Mikrometern. Die zyklische Führung hinterlässt eine wellige Seitenwand (Scalloping). Hauptanwendungen sind mikromechanische Bauelemente (MEMS) und Durchkontaktierungen durch den Wafer (through-silicon vias).
- Kryoätzen: Bei Wafertemperaturen um −100 °C kondensieren Reaktionsprodukte an den kalten Seitenwänden und passivieren sie, während der Ionenbeschuss den Boden freihält. Das ergibt glatte, senkrechte Seitenwände ohne Scalloping. Seit einigen Jahren wird das Verfahren auch beim Ätzen der Speicherlöcher im 3D-NAND-Flash in der Serienfertigung eingesetzt: Bei Tiefen um 10 µm und Aspektverhältnissen über 50:1 erlauben die tiefen Temperaturen Ätzchemien, die bei Raumtemperatur nicht funktionieren, und etwa die doppelte Ätzrate herkömmlicher Dielektrikaprozesse.
- Atomlagengenaues Ätzen (ALE): Aktivierung der Oberfläche und Abtrag werden in getrennte, selbstbegrenzende Halbschritte zerlegt und zyklisch wiederholt. Der Abtrag pro Zyklus liegt im Bereich einer Atomlage, gesteuert wird über die Zyklenzahl. Eingesetzt wird ALE dort, wo einzelne Nanometer

über die Funktion entscheiden, etwa beim Freilegen von Kanälen und beim Ätzen von Gate-Strukturen.

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.

Eingesetzte Ätzgase beim Trockenätzen

Die Prozesse müssen nicht nur mit einer Gasmischung und gleichen Ätzparametern 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. Reale Rezepte bestehen deshalb aus mehreren Schritten mit unterschiedlichen Gasmischungen, Leistungen und Drücken.

Die folgende Tabelle gibt einen Überblick über Gase, die beim Trockenätzen verwendet werden.

Schicht	Prozessgase	Bemerkung
SiO_2, Si_3N_4	CF_4, O_2	F ätzt Si, O_2 entfernt Kohlenstoff (C)
	CHF_3, O_2	CHF_3 wirkt als Polymer, erhöhte Selektivität gegen Si
	CHF_3, CF_4	
	CH_3F, CH_2F_2	verbesserte Selektivität von Si_3N_4 gegenüber SiO_2
	C_2F_6 / SF_6	
	C_3F_8	erhöhte Ätzrate gegenüber CF_4
	C_4F_6, C_4F_8 + O_2, Ar	Standardchemie für tiefe Strukturen (Kontaktlöcher, Speicherlöcher im 3D-NAND); die kohlenstoffreichen Moleküle bilden eine Polymerschicht auf Maske und Seitenwand, O_2 steuert deren Dicke
	BCl_3, Cl_2	keine Kontamination durch Kohlenstoff (C)
	$SiCl_4$, Cl_2 / HCl, O_2 / $SiCl_4$, HCl	
Poly-Si	HBr / Cl_2 / O_2	verbesserte Selektivität gegenüber Fotolack und SiO_2; Standardchemie für Gate-Strukturen
	SF_6	hohe Ätzrate, gute Selektivität gegen SiO_2
	NF_3	hohe Ätzrate, isotrop
	HBr, Cl_2	

monokristallines Si	HBr, NF_3, O_2	höhere Selektivität gegen SiO_2
	BCl_3, Cl_2 / HBr, NF_3	
	SF_6 / C_4F_8 SF_6, O_2	Tiefenätzen: zyklisch im Bosch-Prozess bzw. bei tiefen Temperaturen im Kryoprozess
	Cl_2	isotrope Ätzung
	BCl_3	nur geringe Ätzrate, bindet zusätzlich Restfeuchte und natives Al_2O_3
Al-Legierungen	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
W, TiN, Ti	SF_6 / NF_3	Rückätzen von Wolfram in Kontaktlöchern
	Cl_2, BCl_3	Barrieren- und Metal-Gate-Schichten

Historisch wurden für monokristallines Silicium auch bromhaltige Fluorchlorkohlenwasserstoffe wie CF_3Br verwendet. Sie sind als ozonschichtschädigende Stoffe nach dem Montreal-Protokoll verboten und in der Fertigung durch HBr-Mischungen ersetzt worden.

Metallisierung: warum Kupfer nicht geätzt wird

Die Ätzung von Aluminiumlegierungen hat mit dem Übergang auf Kupferleitbahnen an Bedeutung verloren. Kupfer bildet mit Chlor und Fluor keine bei Prozesstemperatur flüchtigen Verbindungen und lässt sich deshalb praktisch nicht trockenätzen. Stattdessen wird im Damascene-Verfahren zunächst der Graben im Dielektrikum geätzt, anschließend das Kupfer abgeschieden und der Überschuss durch chemisch-mechanisches Polieren entfernt. Aluminium wird weiterhin für Bondpads sowie in der Leistungs- und Analogtechnik strukturiert.

Klimawirkung der Prozessgase

Die perfluorierten Ätz- und Reinigungsgase sind chemisch außerordentlich stabil – genau die Eigenschaft, die sie im Prozess brauchbar macht. In der Atmosphäre werden sie deshalb kaum abgebaut und wirken als sehr starke Treibhausgase. Die Treibhauspotenziale über 100 Jahre liegen nach dem sechsten IPCC-Bericht bei rund 7400 für CF_4, rund 12 400 für C_2F_6, rund 14 600 für CHF_3, rund 17 400 für NF_3 und rund 25 000 für SF_6. CF_4 verbleibt dabei nach heutiger Abschätzung einige zehntausend Jahre in der Atmosphäre.

Fertigungsseitig wird das Abgas jeder Anlage einzeln nachbehandelt (point-of-use abatement), thermisch, katalytisch oder in einem Plasmabrenner. Reaktive Gase wie NF_3 werden dabei zu über 99 % zerlegt, CF_4 jedoch nur teilweise; es

entsteht überdies als Zerlegungsprodukt anderer Fluorverbindungen. Reduziert wird der Beitrag daher vor allem am Prozess selbst: durch Ersatz von C_2F_6 durch NF_3 bei der Kammerreinigung, geringere Gasflüsse, kürzere Ätzzeiten und Verfahren wie das Kryoätzen mit höherer Ätzrate. Regulatorisch sind die Stoffe von der EU-Verordnung über fluorierte Treibhausgase erfasst; zusätzlich wird im Rahmen von REACH eine Beschränkung von PFAS diskutiert, für die Halbleiterfertigung bislang mit Ausnahmen und langen Übergangsfristen.

Sonderfall: Bosch-Prozess (Deep Reactive Ion Etching)

Eine wichtige Weiterentwicklung des reaktiven Ionenätzens ist der Bosch-Prozess (DRIE), der durch zyklischen Wechsel aus Ätz- und Passivierungsschritten extrem hohe Aspektverhältnisse bei nahezu senkrechten Wänden ermöglicht – deutlich mehr, als mit kontinuierlichem Ätzen erreichbar wäre.

Da er vor allem in der Mikrosystemtechnik zur Strukturierung tiefer Gräben und zur Vereinzelung von MEMS-Bauelementen eingesetzt wird, ist er ausführlich im Bereich Mikromechanik beschrieben: DRIE: Deep Reactive Ion Etching.

Anlagentechnik (Reaktortypen)

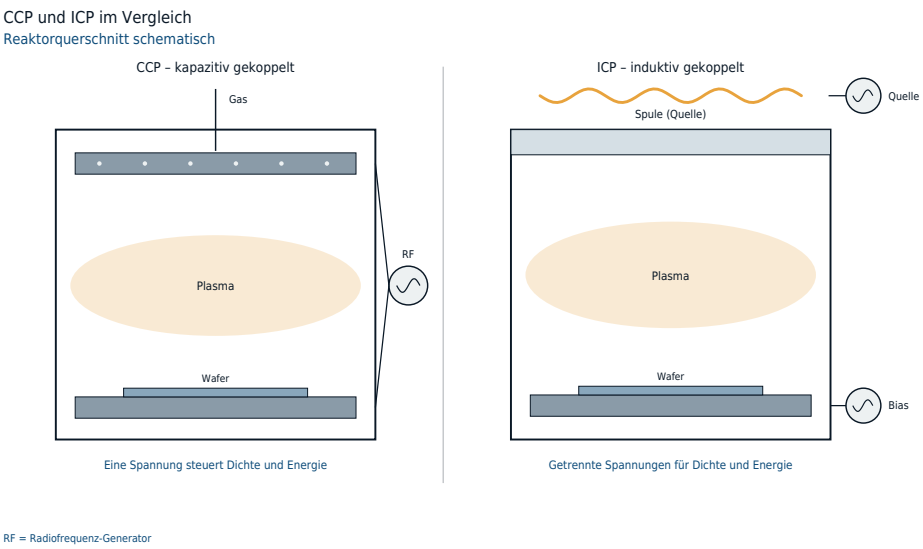
Allgemein: Aufbau einer Trockenätzanlage

Jede Trockenätzanlage besteht aus denselben Grundkomponenten: einer Prozesskammer, einem Gaseinlass – oft als sogenannter Showerhead direkt über dem Wafer, für eine gleichmäßige Gasverteilung – einem Pumpensystem aus Turbomolekularpumpe und vorgeschalteter Drehschieberpumpe für den nötigen Unterdruck, der HF-Einkopplung zur Plasmaerzeugung sowie dem Wafer-Chuck.

Wie diese Komponenten im Detail ausgeführt sind, unterscheidet sich stark zwischen den Reaktortypen – und bestimmt maßgeblich, welche Prozesse eine Anlage beherrscht.

Reaktortypen: CCP und ICP in der Praxis

Wie im Bereich Abscheidung beschrieben, bestimmt bei kapazitiv gekoppelten Reaktoren (CCP) dieselbe Spannung sowohl die Ionendichte als auch die Ionenenergie – beides lässt sich nicht unabhängig einstellen. Induktiv gekoppelte Reaktoren (ICP) trennen diese beiden Größen durch eine separate Bias-Spannung an der Waferelektrode.



Für das Ätzen ist diese Trennung besonders wichtig: Eine hohe Ätzrate braucht viele reaktive Teilchen, also eine hohe Ionendichte, ein sauberes anisotropes Profil dagegen kontrollierte, nicht zu hohe Ionenenergie. Konstruktiv zeigt sich der Unterschied deutlich: CCP-Reaktoren verwenden häufig Silicium- oder SiC-Elektroden als Verbrauchsmaterial, um eine Metallkontamination der Scheibe zu vermeiden; die Kammerwände bestehen meist aus eloxiertem Aluminium, das den aggressiven Fluor- und Chlorradikalen widerstehen muss. Bei ICP-Reaktoren sitzt die Spule außerhalb eines dielektrischen Kammerfensters aus Quarz oder Keramik, durch das das Wechselfeld eindringen kann.

Die erreichbare Ionendichte liegt bei ICP typischerweise um den Faktor 10 bis 100 höher als bei CCP – der Hauptgrund, warum ICP heute der Standard für anspruchsvolle Ätzprozesse mit hohem Aspektverhältnis ist.

Elektronenzyklotronresonanz (ECR)

Bei ECR-Reaktoren erzeugt eine Kombination aus Mikrowelleneinkopplung – meist bei 2,45 GHz – und einem Magnetfeld sehr hohe Plasmadichten bei besonders niedrigem Druck. Die Mikrowellenfrequenz wird dabei genau auf die Zyklotronresonanzfrequenz der Elektronen im Magnetfeld abgestimmt, was eine sehr effiziente Energieübertragung ermöglicht.

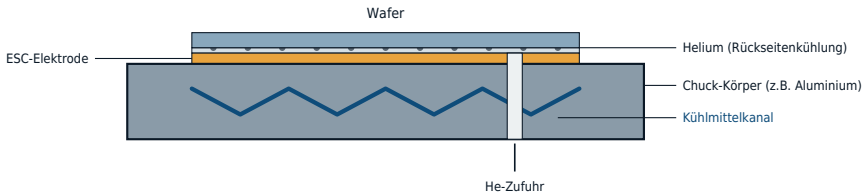
Historisch war ECR ein wichtiger Vorläufer der heutigen ICP-Technik, ist aber wegen der aufwändigeren und teureren Anlagentechnik – zusätzliche Magnetspulen, Mikrowellenquelle statt einfachem RF-Generator – weitgehend von ICP verdrängt worden. Eingesetzt wird es noch in Nischenanwendungen, die besonders niedrige Ionenenergien bei gleichzeitig hoher Dichte erfordern, etwa bei besonders schädigungsempfindlichen Schichten.

Wafer-Chuck: Temperaturkontrolle und elektrostatische Klemmung

Der elektrostatische Chuck (ESC) hält den Wafer während des Prozesses plan und mechanisch stabil, ohne mechanische Klemmen, die den Scheibenrand beschädigen oder abschatten könnten.

Physikalisch wirkt der ESC dabei wie ein Kondensator: Die Gleichspannung an der Elektrode lädt die eine Seite des dünnen Dielektrikums auf, während sich auf der Wafer-Rückseite eine entgegengesetzte Spiegelladung ausbildet. Die daraus resultierende elektrostatische Anziehungskraft presst den Wafer gegen die Oberfläche. Bei einfachen Coulomb-Chucks reicht dafür allein die angelegte Spannung; Johnsen-Rahbek-Chucks nutzen zusätzlich eine leicht leitfähige Dielektrikumsschicht und erreichen dadurch deutlich höhere Klemmkräfte – allerdings mit langsamerem Lösevorgang nach Prozessende, da die Ladung erst abfließen muss.

Wafer-Chuck: Klemmung und Kühlung
Querschnitt schematisch



Elektrostatische Klemmung durch Gleichspannung an der Elektrode

ESC = Electrostatic Chuck (elektrostatischer Chuck)

Da im Vakuum keine Konvektion stattfindet, sorgt eine gezielte Rückseitenkühlung mit Helium-Gas – in feinen Kanälen zwischen Chuck und Waferrückseite eingeleitet – für den nötigen Wärmeübergang. Diese Temperaturkontrolle ist entscheidend: Ätzrate, Selektivität und Profilform hängen alle stark von der Wafertemperatur ab, weshalb moderne Chucks die Temperatur über den gesamten Prozess und über die Scheibe hinweg auf wenige Grad genau regeln.

Ätzen von Verbindungshalbleitern und High-k-Materialien

Warum diese Materialien eine Sonderrolle spielen

Die Gasübersicht im Kapitel Trockenätzverfahren deckt die Standardmaterialien der Siliciumfertigung ab: Silicium, Siliciumdioxid, Siliciumnitrid, Polysilicium und Aluminium. Manche Materialsysteme lassen sich mit denselben Chemien jedoch nicht oder nur schlecht ätzen – entweder weil sie chemisch zu inert sind, oder weil die entstehenden Reaktionsprodukte bei Prozesstemperatur nicht flüchtig genug sind.

III-V-Verbindungshalbleiter (GaAs, InP, GaN)

III-V-Halbleiter wie Galliumarsenid, Indiumphosphid oder Galliumnitrid lassen sich nicht mit Fluorchemie ätzen – die entstehenden Fluoride sind bei Prozesstemperatur nicht flüchtig genug. Stattdessen kommt Chlorchemie zum Einsatz (BCl_3, Cl_2), die mit den Elementen der dritten Hauptgruppe flüchtige Chloride bildet.

Eine Herausforderung ist dabei, dass die verschiedenen Elemente einer Verbindung unterschiedlich schnell reagieren – bei GaAs etwa muss die Chemie so eingestellt werden, dass Gallium und Arsen mit vergleichbarer Rate abgetragen werden. Sonst verarmt die Oberfläche an einem der beiden Elemente, und die Ätzung stoppt oder wird rau.

Siliciumcarbid (SiC)

Siliciumcarbid ist chemisch außerordentlich inert – dieselbe Eigenschaft, die es für Hochtemperatur- und Leistungsanwendungen wertvoll macht, erschwert seine Strukturierung erheblich. Trockenätzen von SiC erfordert Fluorchemie, meist SF_6-basiert, mit deutlich höherer Ionenenergie als bei Silicium, um die starken Si-C-Bindungen aufzubrechen.

Die Ätzraten liegen entsprechend niedriger, und die verwendete Maske erodiert stärker – bei tiefen Strukturen muss die Maskendicke daher großzügiger bemessen werden als bei vergleichbaren Silicium-Prozessen.

High-k-Dielektrika (HfO_2 & Co.)

Hafniumoxid (HfO_2) und verwandte High-k-Materialien, die in modernen Gate-Stacks das klassische Siliciumdioxid als Gate-Dielektrikum ersetzen, bilden mit Fluor- oder Chlorverbindungen nur schwer flüchtige Reaktionsprodukte. Geätzt wird meist mit BCl_3-basierten Chemien, häufig unterstützt durch einen stärkeren physikalischen, also durch Ionenbeschuss geprägten Anteil, um die Reaktionsprodukte mechanisch von der Oberfläche zu entfernen statt sie rein

chemisch abzudampfen.

Die Prozessfenster sind eng, da die extrem dünnen High-k-Schichten – oft nur wenige Nanometer – keine Über- oder Unterätzung erlauben.