

Interner Bericht  
DESY F56-84-01  
Februar 1984

DIE SILIZIUM-HARDWARE DER RECHENMASCHINEN

Gegenwärtiger Stand und Trends

von

H.-J. Stuckenberg

DESY - F56

Eigentum der Property of	<b>DESY</b>	Bibliothek library
Zugang: Accessions:	2 1. MRZ. 1984	
Leihfrist: Loan period:	7	Tage days

**DESY behält sich alle Rechte für den Fall der Schutzrechtserteilung und für die wirtschaftliche Verwertung der in diesem Bericht enthaltenen Informationen vor.**

**DESY reserves all rights for commercial use of information included in this report, especially in case of filing application for or grant of patents.**

**"Die Verantwortung für den Inhalt dieses  
Internen Berichtes liegt ausschließlich beim Verfasser"**

# DIE SILIZIUM-HARDWARE DER RECHENMASCHINEN

## Gegenwärtiger Stand und Trends

H.-J. STUCKENBERG

Deutsches Elektronen-Synchrotron DESY, Hamburg

### 1. Einleitung.

Jeder Rechner (Bild E1) enthält eine Recheneinheit, in die Daten (Operanden) geladen werden, die in einem Speicher vorhanden sind. Ebenfalls im Speicher befindet sich die Rechenvorschrift (Programm), nach der die Operanden miteinander verknüpft werden sollen. Das Ergebnis wird anschließend wieder in den Speicher geschrieben und kann weiter benutzt oder an periphere Elemente (Speicher oder Bildschirme) gegeben werden.

Diese Elemente eines Rechners sind grundsätzlicher Art, sie waren immer vorhanden, solange Menschen rechneten.

Der Rechner der Antike hieß Abakus. Er bestand aus einem Brett mit einer Reihe von Querrinnen, in denen Kugel, Steine oder Münzen hin- und her geschoben wurden. Durch Wertzuordnungen der Querrinnen konnten alle damals üblichen Rechnungsarten durchgeführt werden. Bild E2 zeigt dies am Beispiel einer Rechnung, wie sie ein Kaufmann auf Wunsch eines Kunden durchführt; Adam Riese hat dieses Verfahren genau beschrieben. Bild E3 zeigt neuere Versionen, die auch als Kugelschieber-Rechner bezeichnet werden und heute noch in Asien verwendet werden.

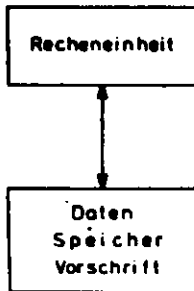
Dies sind natürlich noch keine Rechenmaschinen, es sind manuelle Rechenhilfen. Das Rechnen auf Maschinen begann erst im 17. Jahrhundert mit der Entwicklung einer Rechenuhr. Doch dann dauerte es noch mehr als 300 Jahre, bis die ersten elektromechanischen Rechner, z.B. von Konrad Zuse 1941 gebaut wurden. Bild E4 gibt einen Eindruck von der Größe eines Speichers der Zuse Z4 für 64 Worte. 1946 wurde der erste Rechner mit Röhren (ENIAC) fertiggestellt. Das Bild E5 zeigt die verschiedenen Steckeinheiten mit Röhren, die in den folgenden Jahren verwendet wurden. Erst nach 1950 wurden Transistoren eingesetzt, die die Größe und den Leistungsverbrauch der Rechner enorm verringerte (Bild E6). Nach 1960 begann die Herstellung der ersten integrierten Schaltkreise, die für höhere Komplexität und damit weitere Verkleinerung der Rechenmaschinen sorgten. Heute, gut 20 Jahre später, ist es möglich, eine Millionen Transistoren auf einem Halbleiterkristall, dem Chip, unterzubringen. Wie das möglich ist und was in den nächsten Jahren machbar sein wird, zeigt dieser Bericht.

### 2. Halbleitende Verbindungen.

Halbleiterbauelemente beruhen auf dem Stromfluß in Festkörpern. Allgemein kann man alle Materialien durch ihre Fähigkeit, elektrischen Strom fließen zu lassen, unterscheiden: in Leiter, Halbleiter und Isolatoren. Das meist benutzte Material für halbleitende Verbindungen ist Silizium (Si). Die Struktur reiner Si-Kristalle ist sehr regelmäßig. In ihrer Anordnung entsprechen sie dem Diamantgitter; diese Einkristallstruktur ist ein perfekter Isolator.

In einem Isolator ist das höchste Energieband, das Elektronen enthält, gefüllt (Valenzband), es ist von dem leeren Leitfähigkeitsband durch einen verbotenen Bereich von einigen Elektronenvolt getrennt. Es wird also eine beträchtliche Energie benötigt, um ein Elektron aus dem Valenzband in das Leitfähigkeitsband zu bringen. Ist das Elektron dort angelangt, so ist es frei in seiner Bewegung durch das Kristallgitter. Definieren wir, wie üblich, die Fermigrenze als diejenige Energie, bei der die Wahrscheinlichkeit, daß ein verfügbarer Platz von einem Elektron besetzt ist, gerade 50% beträgt, so ist, da im Isolator die Wahrscheinlichkeit für ein Elektron im Leitfähigkeitsband Null und die im Valenzband 100% ist, die Fermigrenze in der Mitte des verbotenen Bereiches. Dies ist in Bild 1 dargestellt.

In einem reinen Halbleiter sind die Verhältnisse sehr ähnlich. Der bedeutsamste Unterschied ist jedoch die Breite des verbotenen Bereiches, sie ist wesentlich geringer, etwa ein Elektronvolt (vergl. Bild 2). Schon bei Raumtemperatur gelingt es daher einzelnen Elektronen, den verbotenen Bereich zu überspringen und einen Platz im Leitfähigkeitsband zu besetzen. Es entsteht dabei ein Elektron-Lochpaar, d.h. positive und negative Ladungsträger, beide Teile können einen Trägerstrom darstellen. Fließt sowohl negativer als auch positiver Strom, wird er bipolar benannt, fließt nur negativer oder nur positiver Strom, wird er unipolar genannt.



Das Rechenhilfsmittel bis zum hohen Mittelalter: Rechenbrett zum Rechnen auf den Linien. Der Kaufmann legt Rechnung auf seiner Bank mit Rechenpfennigen; der Betrag von 3161 wird gerade zum Kummer des Kunden um 10 erhöht.

### (E1) Elemente des Rechners

### (E2) Rechnen im Mittelalter

#### Adam Riesen.

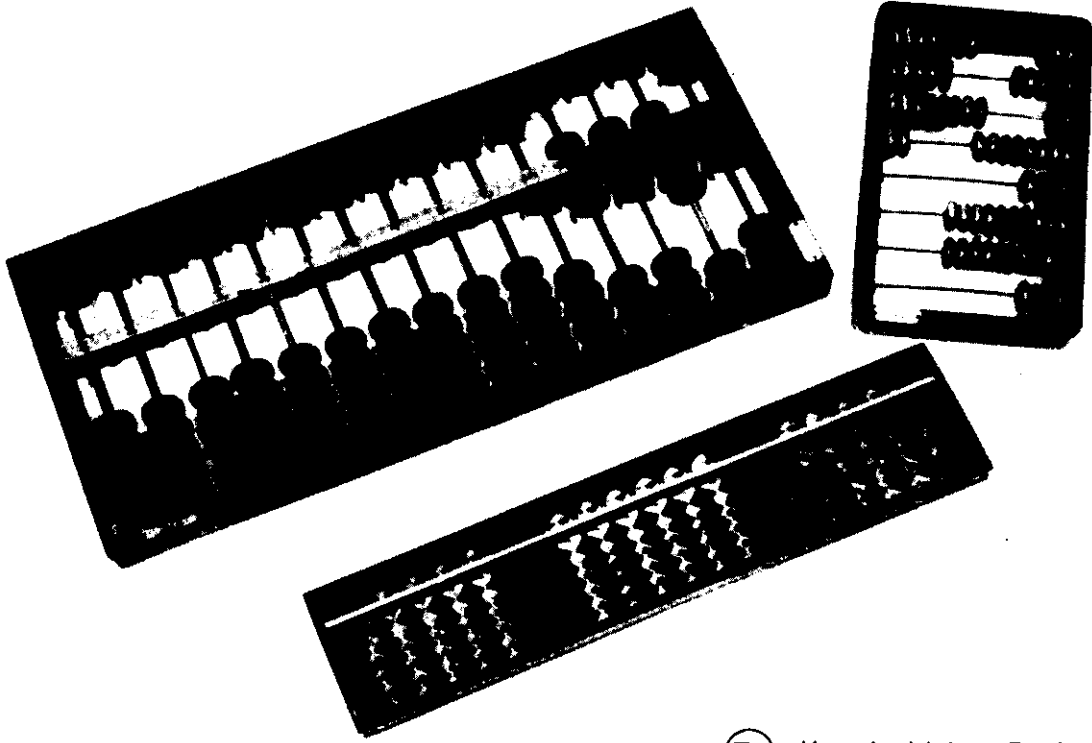
100000	●	6	Hundert tausende
50000	●		Fünffzig tausende
10000	●	5	Zehen tausende
5000	●		Fünff tausende
1000	●	X	Tausende
500	●		Fünff hundert
100	●	3	Hundert
50	●		Fünffzig
10	●	2	Zehen
5	●		Fünff
1	●	1	Eins
$\frac{1}{2}$	●		Ein halbe.

#### Addiren oder Summiren

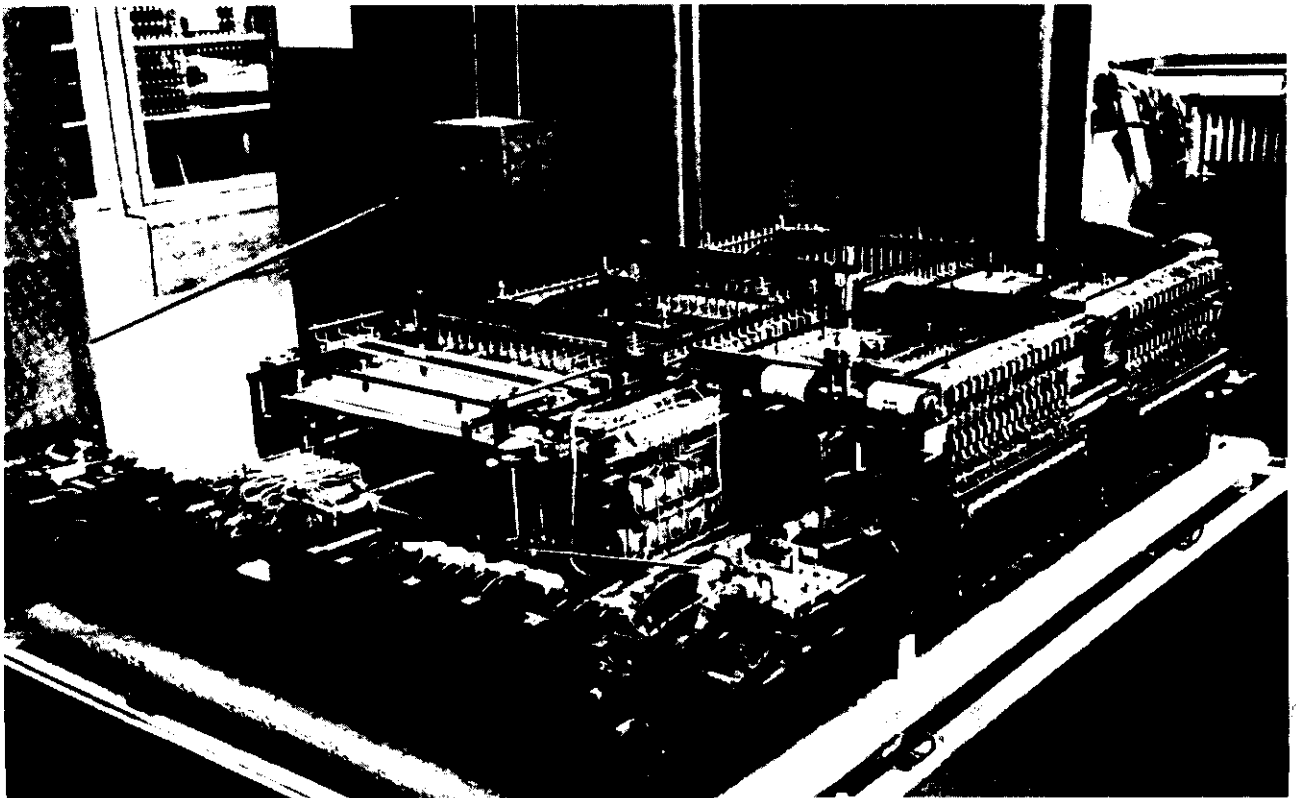
**S**etze zusamen thun / lehret wie man viel vnd mancherley zahlen von gülden/groschen/pfenning vnd hellern in eine summa bringen soll. Thu ihm also: Mache für dich Linien/die theil in so viel feld/als Wärg vorhanden/lege die st. besonder/gro. allein/ dz. vnd heller auch jeglich allein / htr. vnd dz. mach zu gro. was kompt leg zu den gro. Als dann mach die gro. zu st. leg es zu den andern gülden/nach art eines jeglichen Landes.

Auch soltu mercken/wenn fünff dz. auff einer Linien ligen/das du sie auff hebest/vn den fünfften in das nächst spactum darüber legest.

A ih Des



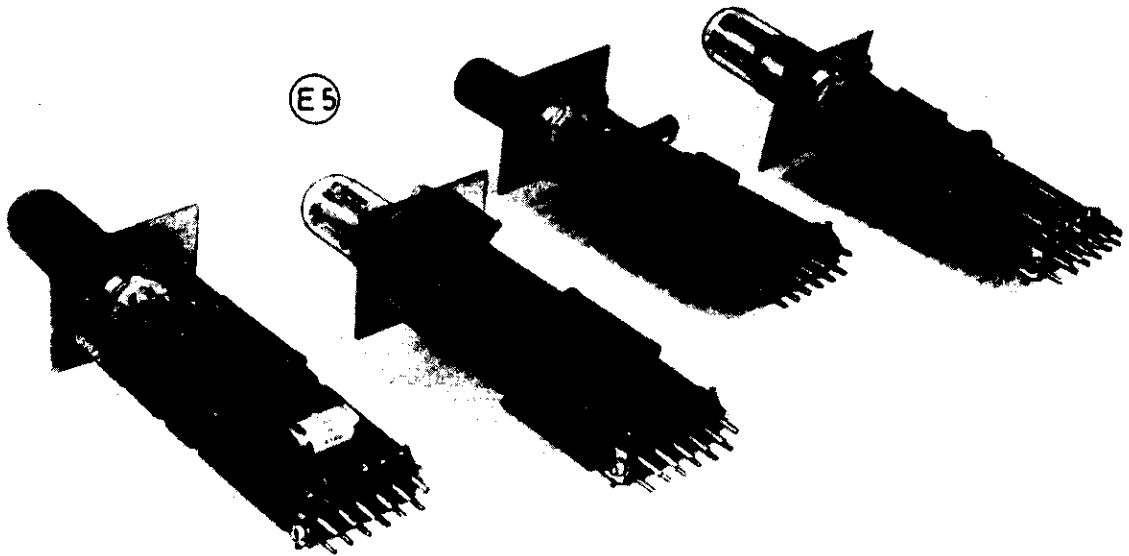
Ⓔ3 Kugelschieber - Rechner



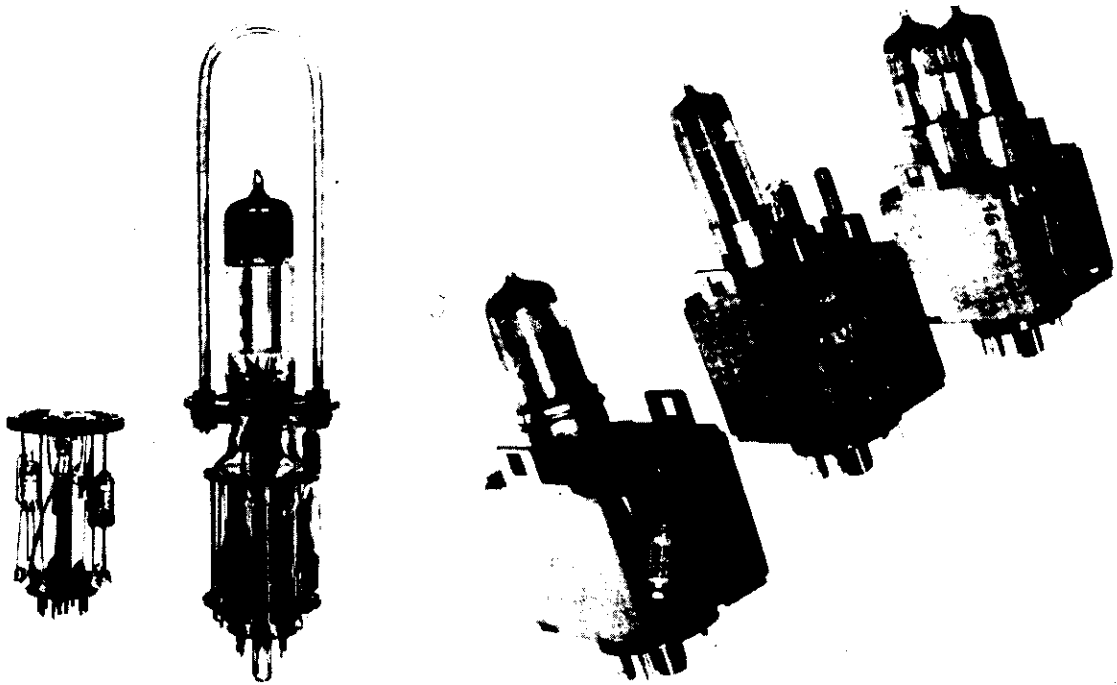
*Mechanisches Speicherwerk der ZUSE Z 4,  
teilausgebaut für 64 Wörter, konstruktiv leicht nach oben  
erweiterbar für 500 Wörter Speicherkapazität.*

Ⓔ4 Relais - Rechner

E5

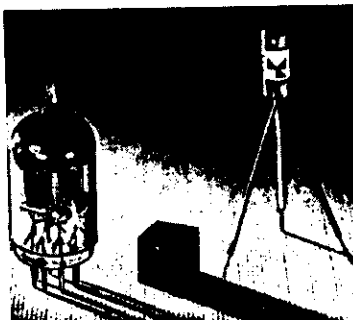


Röhren-Steckeinheiten in Draht-Schaltung (URAL, UdSSR)



Röhren-Steckeinheiten der G 3

Steckeinheiten der PERM (München)  
Impulsgatter, Flip-Flop und Schieberegister-Stelle



E6

Transistor-Röhrenvergleich

Für technische verwendbare Halbleiter, d.h. für Transistoren benötigt man Substanzen, deren Leitfähigkeit durch kontrolliertes Hinzufügen von Verunreinigungen erzeugt werden kann. Die Verunreinigungen, auch Dotierung genannt, bestehen aus bestimmten Atomen, die einige Si-Atome ersetzen. Dabei können zwei verschiedene Materialien zur Dotierung verwendet werden, n-Typ-Material mit negativem Elektronenüberschuß und p-Typ-Material mit positivem Löcherüberschuß. Da Si vierwertig ist, kommen als n-Typ fünfwertige Stoffe, wie Phosphor (P), Arsen (As), Antimon (Sb), als p-Typ dreiwertige Materialien, wie Bor (B), Aluminium (Al), Gallium (Ga) oder Indium (In), in Betracht.

Die typischen Dotierung beträgt etwa 1 Fremdatom auf  $10^6$  bis  $10^8$  Si-Atome.

Gute Leiter, wie Silber (Ag), Kupfer (Cu) und Aluminium (Al), haben eine große Zahl freier Elektronen. Ihr spezifischer Widerstand liegt in der Größenordnung von einigen  $10^{-6} \Omega \cdot \text{cm}$ . Isolatoren, die nur sehr wenige, locker gebundene Elektronen enthalten, haben spezifische Widerstände von etlichen  $10^6 \Omega \cdot \text{cm}$ . Die halbleitenden Materialien liegen dazwischen, reines Si etwa  $60\,000 \Omega \cdot \text{cm}$ . Mit den üblicherweise benutzten Verunreinigungen fallen diese Werte auf etwa  $2 \Omega \cdot \text{cm}$  bei Raumtemperatur.

## 2.1 Bipolare Halbleiter.

Werden Si-Einkristalle mit je einer p- und n-Zone zusammengebracht, entsteht eine Flächendiode. Da diese beiden Gebiete in engem Kontakt sind, ergibt sich eine Diffusion von Ladungsträgern mit der Tendenz, ihre Konzentration, die im p- und n-Bereich zunächst verschieden ist, zu egalisieren (Bild 3). Es wandern also Löcher von der p-Zone zur n-Zone, die Elektronen in umgekehrter Richtung. Durch die Wanderung bleiben Raumladungen zurück, es entsteht eine Raumladungszone mit entsprechendem elektrischen Feld (Bild 4). Die Potentialänderung geschieht in einer schmalen Doppelschicht, der Sperrschicht. In dieser Schicht liegt die Potentialschwelle  $U_t$ , die im Si  $\approx 0.7 \text{ V}$  bei Zimmertemperatur beträgt. Diese Schwelle muß abgebaut werden, damit weiterer Strom fließen kann.

Der Potentialgradient, der durch die Diffusion in der Raumladungszone entsteht, ist in Bild 5 als imaginäre Batterie eingezeichnet. Wird eine externe Batterie an die pn-Zone angeschlossen, bestimmt die Polarität der angelegten Spannung und ihr Einfluß auf die Raumladungszone den Stromfluß. In Bild 6 ist die positive Klemme der externen Batterie an das n-Material, die negative Klemme an das p-Material gelegt. Die freien n-Elektronen werden zum positiven Pol der externen Batterien gezogen, die p-Löcher zum negativen Pol, beide also weg von der Sperrschicht. Die Raumladungszone wird breiter, der Potentialgradient steigt, bis er das Potential der externen Batterie erreicht. Damit endet praktisch der Stromfluß, da keine Potentialdifferenz mehr vorhanden ist; die pn-Schicht ist gesperrt. In Bild 7 ist die positive Klemme der externen Batterie an das p-Material, die negative an das n-Material geschlossen. Jetzt wandern die p-Elektronen zur Batterie, dabei hinterlassen sie neue Löcher. Gleichzeitig kommen Elektronen aus der Batterie in das n-Material und diffundieren zur Sperrschicht. Dadurch wird die Raumladungszone schmaler, und die Energieschwelle nimmt ab. Elektronen aus dem n-Material können dann durch die Sperrschicht hindurch zum p-Material wandern. Dieser Stromfluß bleibt so lange erhalten, wie die äußere Spannung anliegt; die pn-Schicht ist in Leitrichtung vorgespannt.

### 2.1.1 Transistoren.

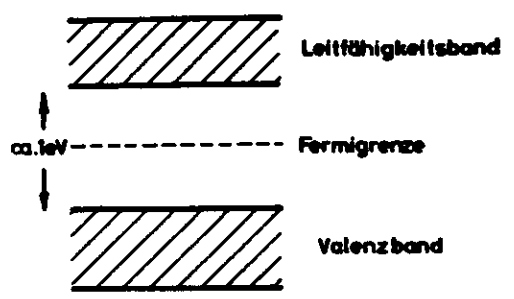
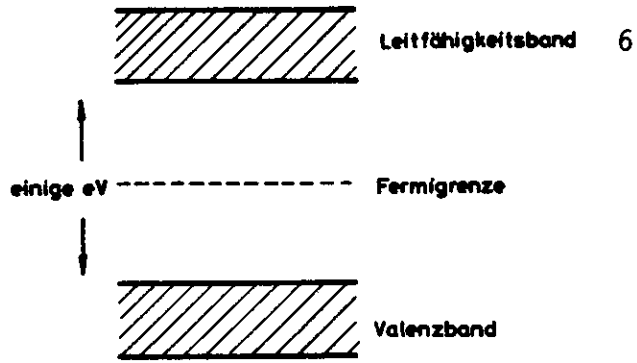
Durch Hinzufügen einer weiteren pn-Schicht (Bild 8) erhält man einen Transistor. Je nach Materialtyp kann man pnp- oder npn-Übergänge erzeugen.

Transistoren arbeiten mit der Injektion, Diffusion und Sammlung von Ladungsträgern. Danach sind die Elektroden als Emitter, Basis und Kollektor bezeichnet.

Die Kollektor-Basis-Strecke ist im allgemeinen gesperrt. Ist auch die Emitter-Basis-Strecke gesperrt, fließt kein Strom durch die Anordnung. Ist die Emitter-Basis-Strecke leitend vorgespannt (also niederohmig), injiziert der Emitter Ladungsträger in die Basis, dort breiten sie sich durch Diffusion aus. Durch die gesperrte hochohmige Kollektor-Basis-Strecke diffundieren die Ladungsträger weiter in die Kollektorzone und werden an der Elektrode gesammelt. Durch Änderung des Potentials an der Basis relativ zum Emitter kann daher der Strom, den der Kollektor sammelt, gesteuert werden (siehe Bild 9). Praktisch ist es so, daß 95 bis 99,5% des Elektronenstroms die Kollektorschicht erreicht. Dieser hohe Prozentsatz ist die Grundlage für die Verstärkungseigenschaft des Transistors.

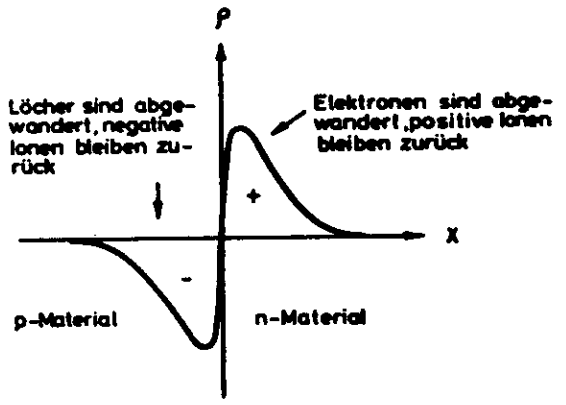
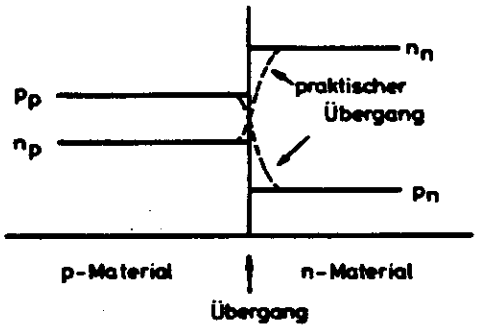
### 2.1.2 Herstellung von bipolaren Transistoren.

Bipolare Transistoren werden heute nach dem Epitaxialplanarverfahren hergestellt. Bei der Epitaxie läßt man auf einem Si-Kristall, der den Kollektor darstellt, eine einkristalline Schicht aufwachsen, die die gleiche



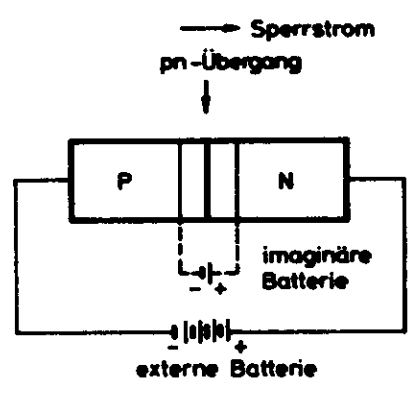
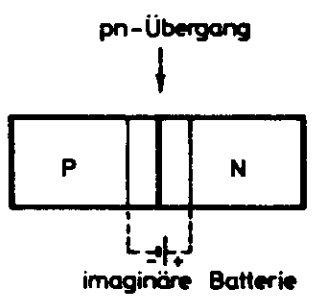
① Energiebänder in Isolatoren

② Energiebänder in Halbleitern



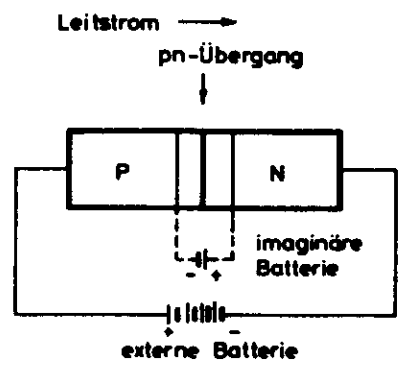
③ Ladungsträger-Konzentrationen im p- und n-Material

④ Aufbau des Raumladungsfeldes am pn-Übergang



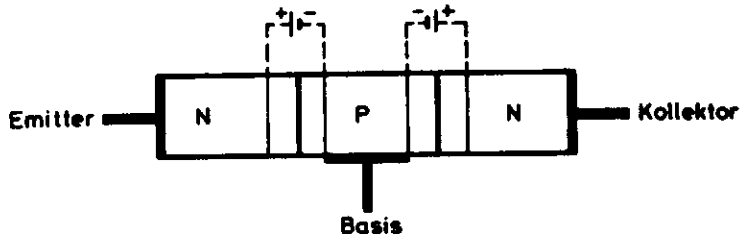
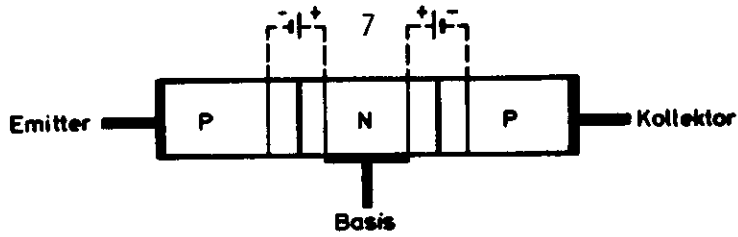
⑤ Potential - Gradient als imaginäre Batterie

⑥ Externe Batterie in Sperrichtung am pn-Übergang

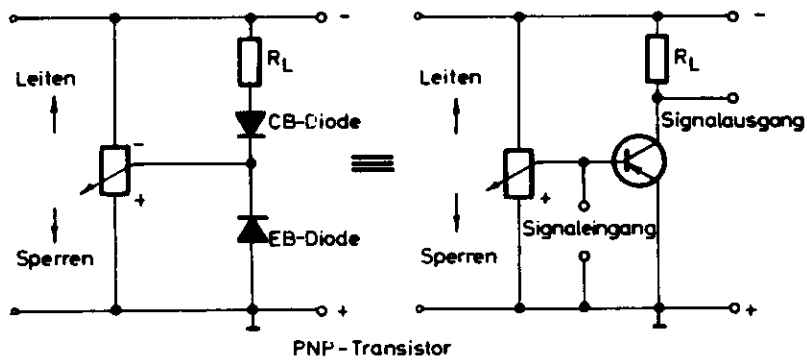
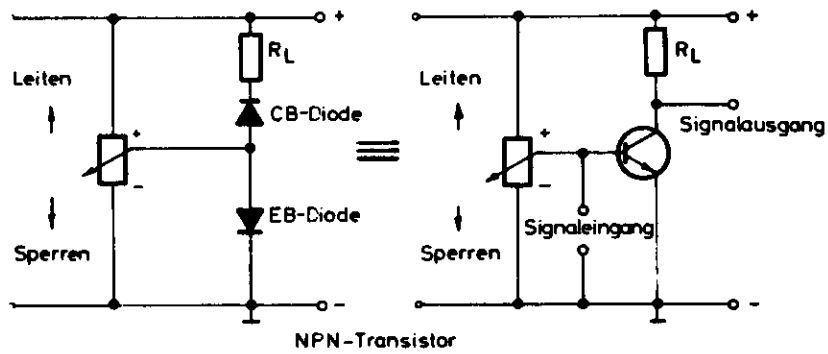


⑦ Externe Batterie in Leitrichtung am pn-Übergang

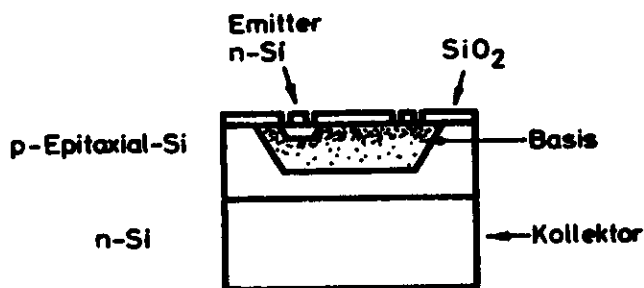




8 Prinzip des Transistors mit 2pn-Übergängen



9 Arbeitspunkteinstellung in Emitterschaltung



10 Planartransistor

kristallografische Orientierung wie der Mutterkristall hat. Diese Schicht stellt die Basiszone dar, in die anschließend der Emitter eindiffundiert wird. Dann werden alle pn-Übergänge durch  $\text{SiO}_2$ -Schichten, die während der Diffusionsvorgänge entstehen, abgedeckt (Bild 10). Das bedeutet vor allem eine lange Lebensdauer für die Halbleiter, da  $\text{SiO}_2$  sehr stabil ist.

## 2.2 Unipolare Halbleiter.

Man kann aber eine gesteuerte Bewegung von Ladungsträgern im Halbleiter auch ganz anders erreichen. Diffundiert man in einen n-Si-Kristall (siehe Bild 11a) drei Zonen zu einer npn-Kombination so, dass die mittlere Schicht nicht das ganze Halbleiterelement durchsetzt, sondern nur den Rand, dann kann das durch diese mittlere Schicht verursachte Feld den Stromfluß zwischen den beiden äußeren Elektroden steuern (siehe Bild 11b). Transistoren dieser Art heißen daher Feldeffekt-Transistoren. Die drei Elektroden heißen hier Quelle, Gate und Senke (Source, Gate, Drain).

Die Strecke zwischen Quelle und Senke wirkt wie ein Halbleiterwiderstand, in dem bei Anlegen einer äußeren Spannung ein Elektronenstrom fließt. Ein Umpolen der Spannung ändert lediglich die Richtung, nicht aber die Stärke des Stromes, dieses Bauelement ist also ein unipolarer Halbleiter. Wird zwischen Quelle S und Gate G eine Spannung in Sperrichtung (Bild 11c) angelegt, beginnt sich die entstehende Sperrschicht mit zunehmender Sperrspannung weiter auszudehnen, bis sie den ganzen Kanal ausfüllt und jeglichen Stromfluß unterbindet. Wird die Gatespannung, bezogen auf die Quelle, gleich oder nahe Null, steigt der Elektronenstrom zunächst linear mit wachsender Spannung zwischen Quelle und Senke. Er erzeugt jedoch im Halbleiterkristall einen Spannungsabfall, der zwischen Gate G und Senke D eine Sperrschicht hervorruft, die sich so weit ausdehnen kann, dass mit zunehmender Spannung kein zusätzlicher Stromfluß möglich ist. Es tritt eine Stromsättigung ein (siehe Bild 11d).

### 2.2.1 MOSFETs.

Eine andere, wesentlich häufiger eingesetzte Art des FETs ist der MOSFET (Metall-Oxyd-Si-FET). Das Ausgangssubstrat ist hier hochohmiges p-Material (Bild 12a), in das zwei getrennte niederohmige n-Regionen, Quelle und Senke, eindiffundiert werden (Bild 12b). Nun wird die Oberfläche mit einer isolierenden Oxydschicht ( $\text{SiO}_2$ ) abgedeckt (Bild 12c), Löcher in das Oxyd geätzt, durch die metallische Kontakte zur Quelle und zur Senke gebracht werden. Die Gateelektrode, die den ganzen Kanal überdeckt, wird auf das Oxyd gedampft, ohne metallisch mit dem Kanal verbunden zu sein (Bild 12d). Ein nennenswerter Strom kann zwischen Quelle und Senke nicht ohne angelegte Gatespannung fließen, da die beiden Übergänge zwei gesperrte Dioden darstellen. Die Metallfläche des Gates bildet mit der  $\text{SiO}_2$ -Schicht und dem darunterliegenden Halbleiterkristall einen Kondensator. Um den Stromfluß zu verstehen, betrachten wir Bild 13. Nehmen wir an, das Gatepotential ist positiv, dann wird durch Influenz im Halbleiterkristall eine negative Ladung erzeugt. Die Halbleiterregion unterhalb der Gateelektrode wird ein n-Kanal, die in der tieferen Oberflächenschicht befindlichen beweglichen Elektronen werden zur Oberfläche gesteuert (Anreicherung von Ladungsträgern). Durch ein elektrisches Feld parallel zur Oberfläche können die Ladungen verschoben werden. Es fließt ein Strom, dessen Stärke durch die Gatespannung gesteuert wird.

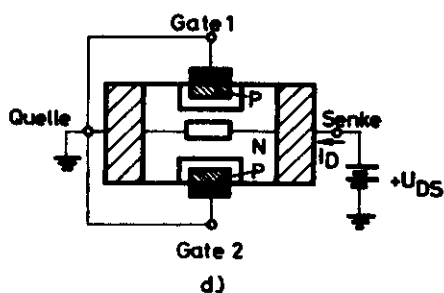
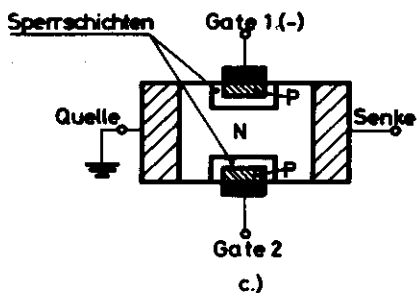
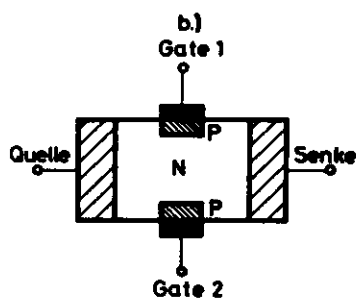
Der beschriebene MOSFET-Typ ist der Anreicherungstyp. Ein Verarmungstyp ist in Bild 14 gezeigt. Hier wird ein Kanal aus n-Material mittleren Widerstands zwischen Quelle und Senke eindiffundiert, so daß jetzt Strom fließen kann, wenn die Gatespannung 0 V beträgt. Für positive Gatespannung arbeitet der MOSFET wie bisher. Bei negativer Spannung am Gatekanal kehrt sich der Prozeß um. Die bisher induzierten Ladungsträger werden aus der Oberfläche vertrieben. Man kann so n- oder auch p-Kanäle erzeugen, entsprechend heißt die Technologie n- oder p-MOS. Da die n-Ladungsträger wesentlich beweglicher als die p-Löcher sind, werden heute meist n-MOS Transistoren verwendet. Eine dritte Technologie kombiniert n- und p-MOS Transistoren parallel geschaltet auf dem Chip. Diese Technik ist besonders leistungsfähig, da, wenn einer die beiden Transistoren in den Strom gefahren wird, der andere herausfährt. Diese Kombination heißt C-MOS (Complementary MOS).

## 3. Integrierte Schaltkreise.

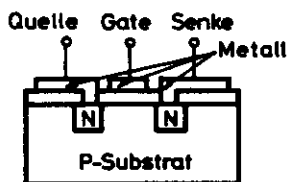
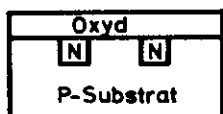
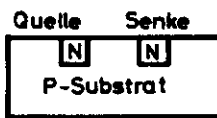
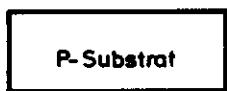
### 3.1 Gründe für die Miniaturisierung.

Die Raumfahrttechnik und unabhängig davon die Computerentwicklung brachten einen Bedarf an elektronischen Bauelementen, die klein, leicht und zuverlässig sein müssen, bei hoher Schaltgeschwindigkeit und möglichst geringer Leistungsaufnahme.

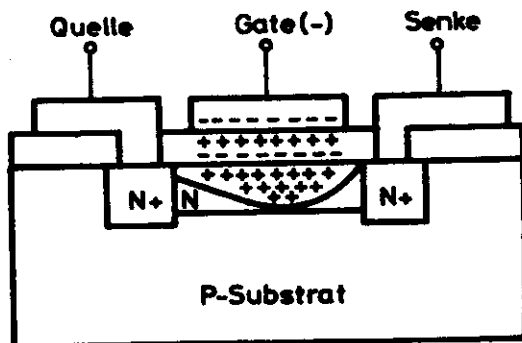
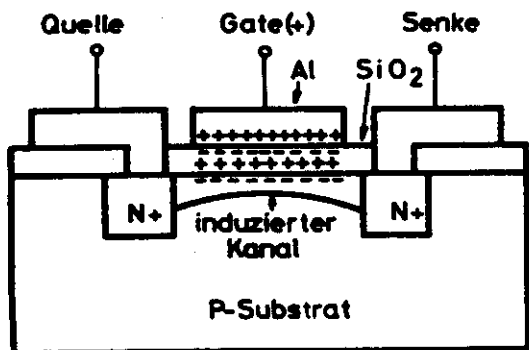
Etwa 1958 begannen die Halbleiterhersteller mit der Entwicklung von integrierten Bausteinen, also Schal-



11 Prinzip eines Feldeffekt-Transistors



12 Aufbau eines n-Kanal MOS-Feldeffekt-Transistors (Anreicherungstyp)



13 Erzeugung eines angereicherten Kanals

14 Erzeugung eines verarmten Kanals

tungen, die eine elektrisch logische Einheit zusammenfassen. Die Transistoren wurden durch verbesserte Herstellungsmethoden ständig kleiner in ihren Abmessungen (Bild 15).

Der Ge-Transistoren erzeugte noch 11mW Wärme, die jetzigen nur noch 0.0001 mW. Dieser Fortschritt wurde durch Forderungen aus der Anwendung erzwungen. Die Entwicklung integrierter Schaltungen wurde vor allem durch die hohen Zuverlässigkeitsanforderungen forciert, die von der Raumfahrt gestellt wurden. Die Gesamtzuverlässigkeit einer konventionellen Schaltung wird von den einzelnen Bauelementen, ihrer Anzahl und der Zahl der gelöteten Zwischenverbindungen bestimmt, die einer integrierten Schaltung kann mit der Zuverlässigkeit eines Einzeltransistors verglichen werden, weil zum Herstellen einer integrierten Schaltung dieselben Verfahren mit nur wenig mehr Prozessschritten angewandt werden wie bei der Transistorfertigung. Die Zahl der Zwischenverbindungen wird stark reduziert. Schließlich ist es auch wichtig, daß integrierte Schaltungen sich sehr wirtschaftlich herstellen lassen, da einige Tausend oder Hunderttausend Schaltungen gleichzeitig in einem Fertigungsschritt bearbeitet werden können.

Unter einer integrierten Schaltung versteht man im allgemeinen eine monolithische Schaltung, deren Herstellung im folgenden kurz beschrieben wird. Als monolithisch bezeichnet man integrierte Schaltungen, die auf einem einzigen Kristallplättchen aus einkristallinem Silizium untergebracht sind. Alle zur Schaltung gehörenden Bauelemente sind auf oder in dem gemeinsamen Kristall. Sie sind durch isoliert aufgebrachte, metallische Leitbahnen untereinander verbunden.

Die Anordnung mehrerer Elemente in einem gemeinsamen Kristall von endlichem Widerstand erfordert Maßnahmen zur gegenseitigen elektrischen Isolierung. Dies wird z.B. mit Hilfe gesperrter pn-Übergänge erreicht. Auf den Ausgangskristall, der in fast allen Fällen p-leitend ist und als Substrat bezeichnet wird, bringt man eine n-leitende Schicht auf, die durch rahmenförmige p-Zonen in getrennte Inseln aufgeteilt wird (Bild 16). Diese Inseln sind innerhalb des Kristalls von pn-Übergängen umgeben, die gesperrt werden (Bild 17). Die Ausdehnung der pn-Übergänge sowie ihre Kapazität beeinflussen die benötigte Chipfläche sowie die Kapazitäten der Elemente gegeneinander und damit die Laufzeiten innerhalb des Kristalls.

Der Aufbau der Transistoren innerhalb der Inseln ist ähnlich der im diskreten Element, nur die Kollektoren müssen jetzt in der Ebene der Basen und Ermitter eindiffundiert werden (Bild 18). Damit jetzt die Kollektorstromwege nicht zu lang und damit zu hochohmig werden, werden niederohmige vergrabene Schichten eindiffundiert (Bild 19).

### 3.2 Herstellung der integrierten Schaltkreise.

Integrierte Schaltkreise werden nach dem von der Transistorfertigung bekannten Planarverfahren hergestellt. Der Halbleiterkristall wird zunächst mit einer Oxydschicht aus  $\text{SiO}_2$  abgedeckt, die als Maske für die zur Dotierung verwendeten Stoffe eingesetzt wird. An den Stellen, wo der Kristall dotiert werden soll, ätzt man fotolithografisch Öffnungen in das Oxyd, durch die der Dotierungsstoff während der Diffusion in den darunterliegenden Si-Kristall einwandern kann. Bild 20 zeigt diesen Vorgang. Durch wiederholtes Abdecken mit  $\text{SiO}_2$  und Ätzung neuer Fenster können im Si-Kristall übereinanderliegende Schichten von p- und n-Zonen hergestellt werden, die die gewünschte Doptierung haben.

Um einen Transistor zu erzeugen, müssen alle p- und n-Zonen innerhalb eines einheitlich orientierten Kristallgitters angeordnet sein. Diese Einkristalle werden meist im Tiegel-Ziehverfahren hergestellt, anschliessend schneidet man die Scheiben und poliert sie. Von der Güte dieser Schnitte hängt der reproduzierbare Verlauf der Diffusions- und Ätzprozesse ab.

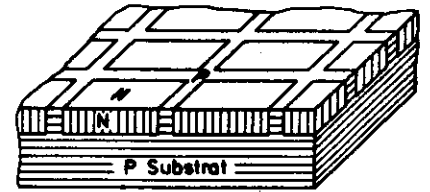
Das bereits erwähnte fotolithografische Verfahren, mit dem die verschiedenen Zonen und Leiterbahnen im Kristall hergestellt werden, ist eigentlich ein Fotoätzverfahren, weil der verwendete Fotolack als Maske zum Ätzen der Öffnung im Oxyd dient. Das prinzipielle Verfahren ist in Bild 21 dargestellt.

Auf die  $\text{SiO}_2$ -Schicht wird erst der Fotolack aufgebracht, anschliessend wird die Scheibe gegenüber der fest eingespannten Fotomaske justiert. Nun wird mit ultraviolettem Licht belichtet, dann die Schicht entwickelt. Dabei werden die belichteten Lackflächen bleiben, die unbelichteten jedoch beseitigt. Nun werden die Fenster in der Oxydschicht weggeätzt, anschliessend der Restlack entfernt. Sämtliche beschriebenen Verfahren werden in einem Raum mit Gelbbeleuchtung ausgeführt, gegen die der Fotolack unempfindlich ist. Die Qualität der Fotomasken hat entscheidenden Einfluß auf die Ausbeute und die Eigenschaften der Schaltungen. Die Originalzeichnung der Masken wird daher in starker Vergrößerung 500 : 1 bis 200 : 1 gefertigt und geprüft, ehe sie verkleinert wird.

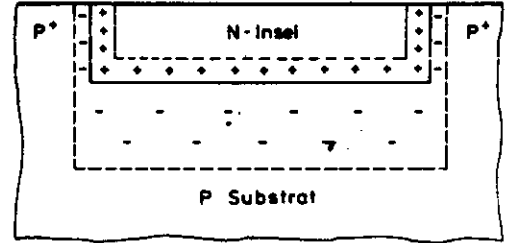
15 Verkleinerung der Transistorabmessungen und der Verlustleistung

1952 Ge-Transistor 2,5 x 4 mm 10 mm<sup>2</sup> 11mW  
 1962 Si-Transistor 0,35x 0,35mm 0,1  
 1972 Si-Transistor 30 x 35 μm 0,001  
 1982 Si-Transistor 3 x 3,5 μm 0,00001 0,1 μW

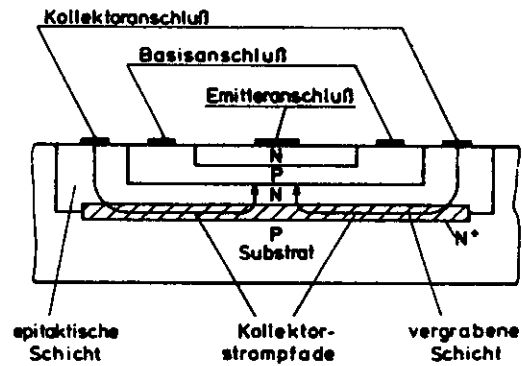
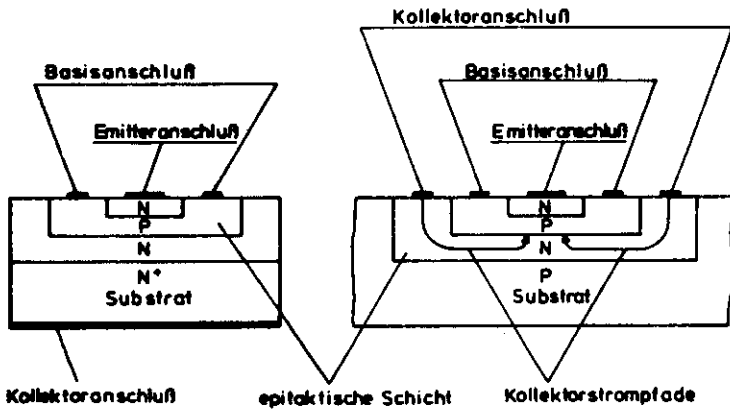
Die Fläche wurde 10<sup>6</sup>: 1 verkleinert  
 Die Verlustleistung wurde 10<sup>5</sup>: 1 verkleinert



16 Inselbildung im Si-Kristall

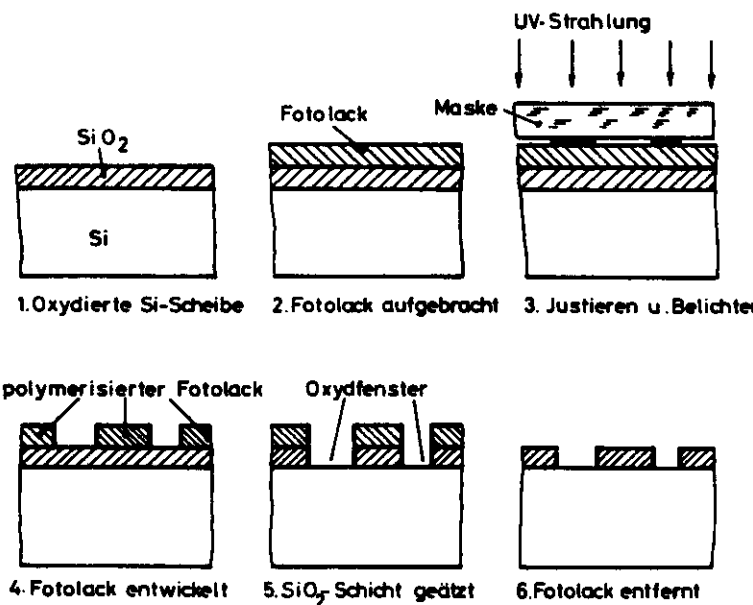
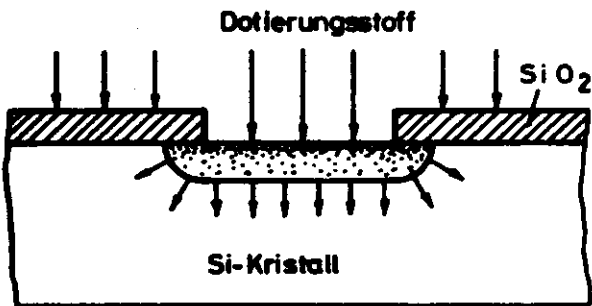


17 Sperrn der Insel



18 Einzelner Planartransistor Transistor in integrierter Form

19 Vergrabene Schicht



20 Prinzip des Planarverfahrens

21 Schematische Darstellung des fotolithografischen Prozesses

### 3.2.1 Fertigungsgang eines IC's.

Als Ausgangsmaterial für integrierte Halbleiterschaltungen verwendet man p-leitende Siliziumkristallscheiben die z.B. einen Durchmesser von 30 bis 125 mm, eine Dicke von 0,25 mm und einen spezifischen Widerstand von 2 bis 5  $\Omega \cdot \text{cm}$  haben. Nach einer ersten Oxydation werden in die entstandene  $\text{SiO}_2$ -Schicht (0,7  $\mu\text{m}$  dick) mittels Fotolithografie Fenster für die vergrabenen Schichten geätzt. Durch diese Fenster läßt man Arsen 1 bis 2  $\mu\text{m}$  tief in das Substrat diffundieren (Bild 22a), woraus dann im Verlauf weiterer Wärmeprozesse die n-leitenden vergrabenen Schichten in ihrer endgültigen Form entstehen. Durch erneute Oxydation und Fotolithografie erhält man rahmenförmig angeordnete Gräben, in die man eine bestimmte Menge Bor niederschlägt und ein wenig eindiffundieren läßt (Bild 22b), so daß stark dotierte  $p^+$ -Zonen entstehen, die nachher bei der Isolierdiffusion als Diffusionsquelle wirken.

Nun läßt man eine etwa 10  $\mu\text{m}$  dicke Schicht epitaktisch aufwachsen (Bild 22c), die mit Antimon schwach n-dotiert wird (0,3  $\Omega \cdot \text{cm}$ ). Aus dieser Schicht bilden sich nachher die isolierten n-Inseln. Nach Oxydation und einem fotolithografischen Prozeß mit derselben Maske wie in Bild 22b wird wiederum Bor niedergeschlagen. Jetzt erfolgt die eigentliche Isolierdiffusion (2 bis 3 Stunden bei 1200 Grad C). Dabei vereinigen sich die rahmenförmigen p-Strukturen oberhalb und unterhalb der epitaktischen Schicht und trennen diese in einzelne n-leitende Inseln auf (Bild 22d). Jede dieser Inseln ist im Kristall allseitig von p-leitendem Material umgeben. Während der inzwischen abgelaufenen Wärmeprozesse, hauptsächlich aber während der am längsten dauernden Isolierdiffusion, hat sich die vergrabene Schicht ( $n, n^+$ ) immer weiter aufgebläht und nahezu ihre endgültige Größe erreicht.

Schon während der Isolierdiffusion hat sich die Siliziumoxydschicht über den p-dotierten Rahmen bereits wieder geschlossen, und mit Hilfe einer dritten Fotomaske werden die Oxydfenster für die Diffusion der p-leitenden Basiszonen und Widerstandsbahnen geätzt. Wieder deponiert man Bor und läßt es anschließend so weit in die epitaktische Schicht diffundieren, daß in einer Tiefe von 3,5  $\mu\text{m}$  ein pn-Übergang entsteht (Bild 22e).

In die während der Basisdiffusion gebildete Oxydschicht werden fotolithografisch neue Fenster für die Emitterzonen, die  $n^+$ -Zonen in den Kollektorgebieten (zur Realisierung sperrschichtfreier Kollektoranschlüsse) und für eventuelle Leitbahnbrücken geätzt. Die erforderliche n-Leitfähigkeit wird mittels Phosphordiffusion von hoher Oberflächenkonzentration erhalten; die Diffusionsfront läßt man so weit eindringen, daß die Emitter-Basis-Übergänge der Transistoren in einer Tiefe von 2,6  $\mu\text{m}$  entstehen (Bild 22f). Gleichzeitig mit der Phosphordiffusion wachsen die Oxydfenster wieder zu (Phosphorglas).

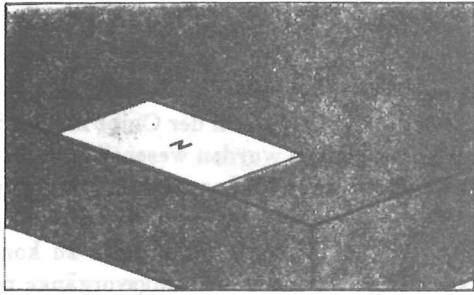
Nun wird die Kristallscheibe von der Rückseite (Unterseite) her abgeätzt, so daß alle während der vorhergehenden Prozesse von unten eingedrungenen Stoffe (Bor, Phosphor,  $\text{SiO}_2$ -Schichten) beseitigt sind. Gleichzeitig bringt man mit dieser Ätzung die Kristallscheibe auf eine günstige Dicke (0,2 mm) für das spätere Ritzen und Brechen. Dann wird auf die Rückseite eine bestimmte Menge Gold aufgedampft und anschließend eindiffundiert.

Mit einer fünften Fotomaske werden nun Kontaktierungslöcher in der Oxydschicht geöffnet, um Transistorzonen und Widerstandsbahnen sowie das p-leitende Substrat und gegebenenfalls die n-leitenden Inseln (falls diese nicht schon als Kollektor verwendet werden) anschließen zu können. Nach Bedampfung der ganzen Kristalloberfläche mit einer ca. 1  $\mu\text{m}$  dicken Aluminiumschicht wird mit Hilfe eines sechsten und letzten fotolithografischen Prozesses das überflüssige Aluminium wieder entfernt, nur Leitbahnen und Anschlußflecken bleiben stehen. Mit einer kurzen Erwärmung auf etwa 550° C, bei der das Aluminium mit dem Silizium an den Anschlußstellen einen guten Kontakt bildet, ist die Bearbeitung der Kristallscheiben als Ganzes abgeschlossen (Bild 23).

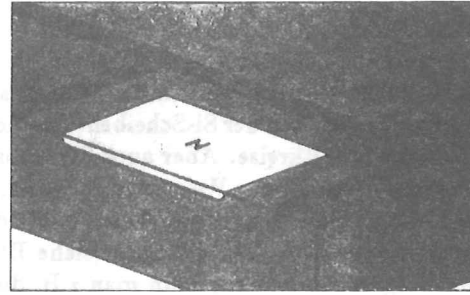
Vor dem Zerteilen der Kristallscheiben werden die einzelnen Schaltungen durch Aufsetzen von Meßspitzen (30). Ans vorgemessen. Anschließend werden die Scheiben mit Diamanten rasterförmig geritzt und gebrochen. Die beim Vormessen als funktionsuntüchtig gekennzeichneten Halbleiterplättchen bzw. -schaltungen werden ausgeschieden. Die guten werden auf den Boden eines Gehäuses gelötet und mit Hilfe eines Thermokompressionsverfahrens kontaktiert. Nach vakuumdichtem Verschuß des Gehäuses und einer Endkontrolle sind die integrierten Halbleiterschaltungen fertig.

### 3.2.2 Verbesserungen der Methoden, Neuerungen.

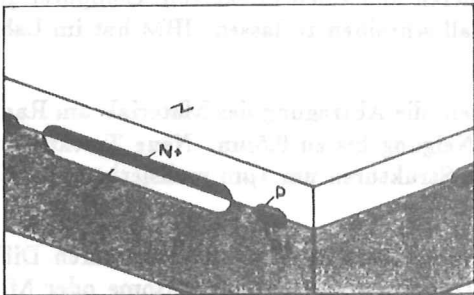
Der eben geschilderte Ablauf ist in den vergangenen 10 Jahren an manchen Punkten verbessert worden. So sind die meisten Abläufe heute weitgehend automatisiert (Bild 24). Der Stromlaufplan des Schaltkreises wird entworfen, ebenso das Chip-layout. Dann wird über Rechner das logische Verhalten der Schaltung berechnet sowie das Zeitverhalten der Signale in der Schaltung. Erst wenn die Funktion sowie das Timing innerhalb



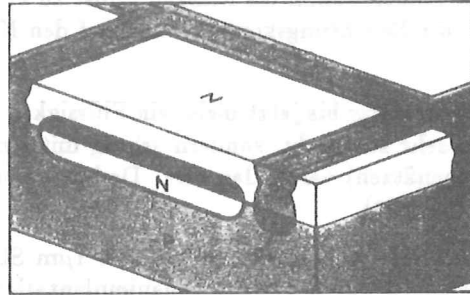
a) Diffundieren von Arsen in das P-leitende Substrat zur Erzeugung N-leitender vergrabener Schichten (buried layers) mit niedrigem Widerstand



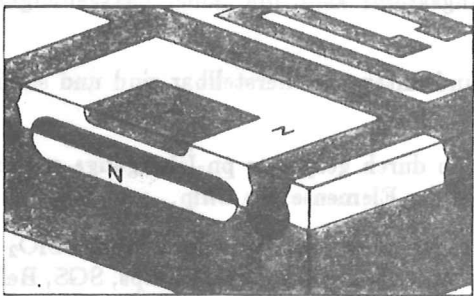
b) Diffundieren rahmenförmiger P-Zonen (Bor) zum Zwecke der Inselbildung



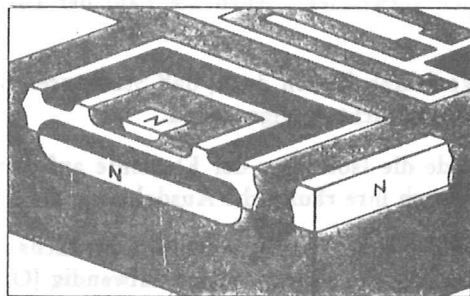
c) Aufwachsen einer mit Antimon schwach dotierten N-leitenden epitaktischen Schicht (Kollektorzonen)



d) Rahmenförmige P-Dotierung mit Bor (gestrichelt); anschließend Isolierdiffusion, d. h. Bildung N-leitender Inseln (dick ausgezogen)



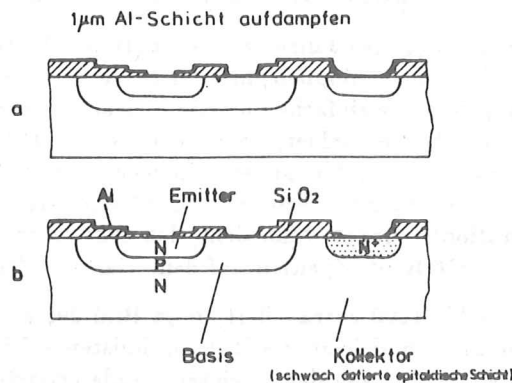
e) Diffundieren von Bor für die P-leitenden Basiszonen und die Widerstandsbahnen



f) Diffundieren von Phosphor für die N-leitenden Emitterzonen und für die N-Zonen zur sperrschichtfreien Kontaktierung hochohmiger N-Zonen

22 Fertigungsgang einer integrierten Halbleiterschaltung

Das Substrat ist, verglichen mit der Dicke der aktiven Schichten, in Wirklichkeit nicht so dünn wie hier abgebildet, sondern 25mal so dick (250  $\mu\text{m}$ ) wie die epitaktische Schicht (10  $\mu\text{m}$ ). Die Siliziumoxydschicht ist in dieser Darstellung weggelassen.



Nach Belichten und Ätzen wird überflüssiges Al entfernt. Kontakte sind frei.

23 Herstellung der Leitbahnen

des Schaltkreises optimiert wurden, werden die Masken sowie der Halbleiter selbst hergestellt und automatisch geprüft. Dann erfolgt das Trennen der Si-Scheiben, das Kontaktieren, das Verpacken der Chips im Gehäuse sowie die Prüfung der fertigen Schaltkreise. Aber auch die technologischen Abläufe wurden wesentlich verbessert. Die Genauigkeit, die auf die beschriebene Weise erreichbar ist, liegt bei Strukturen bis herunter zu etwa  $2\mu\text{m}$ , bedingt durch die Art der Lithographie, des Ätzverfahrens sowie der Diffusionsprozesse. Mitte der siebziger Jahre waren Leiterbreiten von  $6\text{--}8\mu\text{m}$  typisch, heute  $2\text{--}3\mu\text{m}$  (siehe Bild 25). Um zu kleineren Strukturen zu kommen und damit zu höheren Integrationsdichten, kann man z.B. die Wellenlänge für die Belichtungsvorgänge reduzieren, denn das bisher verwendete UV-Licht hat  $0,35\text{--}0,4\mu\text{m}$  Wellenlänge. Beugungen und Interferenzerscheinungen begrenzen die praktische Anwendungen bis etwa  $2\mu\text{m}$ , bei weiterer Verfeinerung kommen Elektronenstrahl- und Röntgenlithographie zum Einsatz, mit denen man im Labor Strukturen von unter  $1\mu\text{m}$  erreicht. Die Entwicklung geht jedoch dahin, die Masken selbst zu eliminieren und einen direkt vom Computer gesteuerten Elektronenstrahl die Belichtungskonfiguration auf den Kristall schreiben zu lassen. IBM hat im Labor bereits Erfolge.

Das Ätzverfahren war bis jetzt meist ein Flüssigkeitsätzen, die Abtragung des Materials am Rand geschah dadurch nicht so sehr senkrecht, sondern schräg mit einer Neigung bis zu  $0,5\mu\text{m}$ . Neue Trockenätzverfahren (Plasmaätzen, Ionenätzen) vermeiden dies. Dadurch werden Strukturen um  $1\mu\text{m}$  realisierbar, das Unterätzen wird vermieden (Bild 26).

Die Dotierung erfolgt beim Herstellen von  $1\mu\text{m}$  Strukturen auch meist nicht mehr durch Diffusion aus der Gasphase, sondern durch gezielte Ionenimplantation. Dabei werden ionisierte Atome oder Moleküle im elektrischen Feld beschleunigt und in den Festkörper geschossen (Bild 27). Die Eindringtiefen sind abhängig von ihrer Masse sowie von der Ionenenergie, bei  $10\text{ keV}$  etwa  $0,01$  bis  $0,1\mu\text{m}$ , bei  $100\text{ keV}$  bis zu  $1\mu\text{m}$ . Die Implantation ist ein Niedertemperaturprozess im Gegensatz zur Diffusion, man kann daher andere einfachere Maskenmaterialien verwenden. Auch die Zahl der Dotierungsatome kann mit hoher Ortsgenauigkeit dosiert werden.

Alle diese Verfahren führen dazu, daß Strukturen um und unter  $1\mu\text{m}$  herstellbar sind und somit für eine sehr hohe Integrationsdichte geeignet sind.

Anfangs wurde die Isolierung der Elemente auf dem Chip durch gesperrte pn-Übergänge erwähnt. Auch diese begrenzen durch ihre räumliche Ausdehnung die Anzahl der Elemente pro Chip.

Verschiedene Hersteller verwenden unterschiedliche neuere Prozesse, z.B. die Isolation durch  $\text{SiO}_2$  verringert die Chipfläche um  $50\%$  ist aber technisch aufwendig (Oxydwallisolierung Fairchild, Philips, SGS, Bell) oder V-förmige Gräben, die mit  $\text{SiO}_2$  beschichtet, eventuell mit polykristallinem Si wieder aufgefüllt werden (Motorola). Auch hier konnte eine deutliche Verringerung der benutzten Fläche und damit eine entsprechende Erhöhung der Integrationsdichte erreicht werden. Was ist als Integrationsdichte denn überhaupt erreichbar? Neuere Überlegungen, die die Chipgröße, Signallaufzeiten sowie ein Geschwindigkeits-Leistungsprodukt berücksichtigen, kommen auf minimale Strukturbreiten von etwa  $0,1\mu\text{m}$ . Wenn man solche Strukturen mit einer denkbaren Chipfläche von  $1000\text{ mm}^2$  ( $3,3 \times 3,3\text{ cm}$ ) kombiniert, könnte man  $10^7$  bis  $10^9$  Elemente, also bis zu 1 Milliarde Transistoren auf einem Chip unterbringen. Solch große Chips würden aber Laufzeitprobleme haben, d.h. keine sehr kurzen Schaltzeiten. Entweder heißt das deutlich kleinere Chips oder kleinere Inseln innerhalb des Chips, die jeweils nur noch intern synchron getaktet arbeiten, untereinander aber asynchron miteinander kommunizieren.

### 3.3 Integrationsgrenzen.

Die Größtintegration strebt also einen asymptotischen Grenzwert zu, aber welchem?

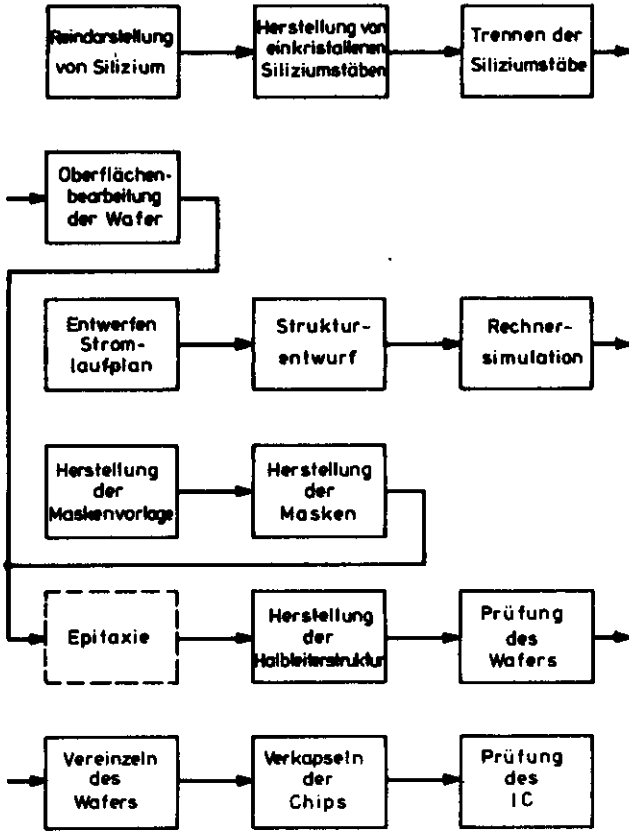
Wenn man die Entwicklung der vergangenen Jahre betrachtet (Bild 28), erkennt man, daß sich die Zahl der Transistorfunktionen je Chip fast jedes Jahr verdoppelt, in zehn Jahren also vertausendfacht. Die Entwicklung hat der Komplexität auch Namen gegeben, die einfache Logik, d.h. Gates und FF's nennt man SSI (Small Scale Integration), die zusammengesetzte Logik, d.h. Zähler, Register mit bis zu  $10^3$  Funktionen heißt MSI (Medium Scale Integration), bis zu etwa  $10^6$  Funktionen, d.h. größere Speicher oder Mikroprozessoren heißt LSI (Large Scale Integration), die Integration kommender Großspeicher oder kompletter Rechnersysteme auf einem Chip wird VLSI (Very Large Scale Integration) genannt. Man sieht, daß 1984 schon einige Mega-Funktionen erreicht sein müßten, tatsächlich gibt es z.Zt.  $256\text{ KBit}$ -Speicher auf dem Markt,  $1\text{ MBit}$ -Chips sind im Labor erreicht.

Diese Entwicklung für die Jahre bis 1990 extrapoliert, zeigt Bild 29, wo man die bereits erwähnten  $10^9$  Funktionen/Chip tatsächlich erwartet. Da sich aber selbst Optimisten solche Integrationsgrade nur schwer vorstellen können, wird von vielen ein Abknicken der gezeigten Grade erwartet. Es ist jedoch bedeutungslos,

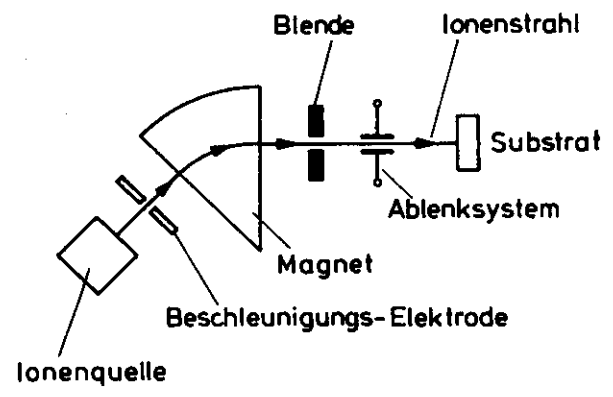


25 Verkleinerung der Strukturen, verwendetes Licht

	Struktur	Litographie
1965	10 $\mu\text{m}$	0,35 - 0,4 $\mu\text{m}$ UV
1975	6-8 $\mu\text{m}$	0,2 - 0,3 $\mu\text{m}$ Deep UV
1985	1-3 $\mu\text{m}$	Elektronenstrahl



27 Prinzipdarstellung einer Ionenimplantationsapparatur



Ionen	Ionenenergie in KeV			Reichweite in $\mu\text{m}$
	10	100	1000	
P	B	0,04	0,4	2,5
	Al	0,02	0,1	1,3
	Ga	0,01	0,05	0,55
N	P	0,02	0,1	1,2
	As	0,01	0,06	0,6
	Sb	0,007	0,05	0,35

24 Automatisierter Ablauf der IC-Herstellung

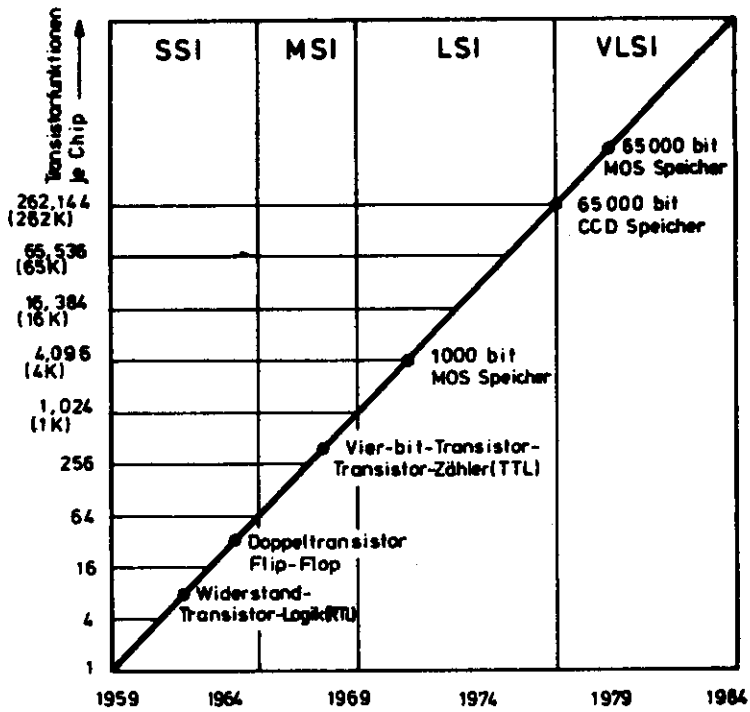


Plasma-Ätzen von Aluminium-Bahnen. Die jeweiligen Strukturbreiten sind im Bild angegeben. Wie man sieht, sind die 1- $\mu\text{m}$ -Strukturen nicht mehr darstellbar: die vier Bahnen sind zu einer breiten verschmolzen. Hier zeigt sich die Grenze der Fotolithografie

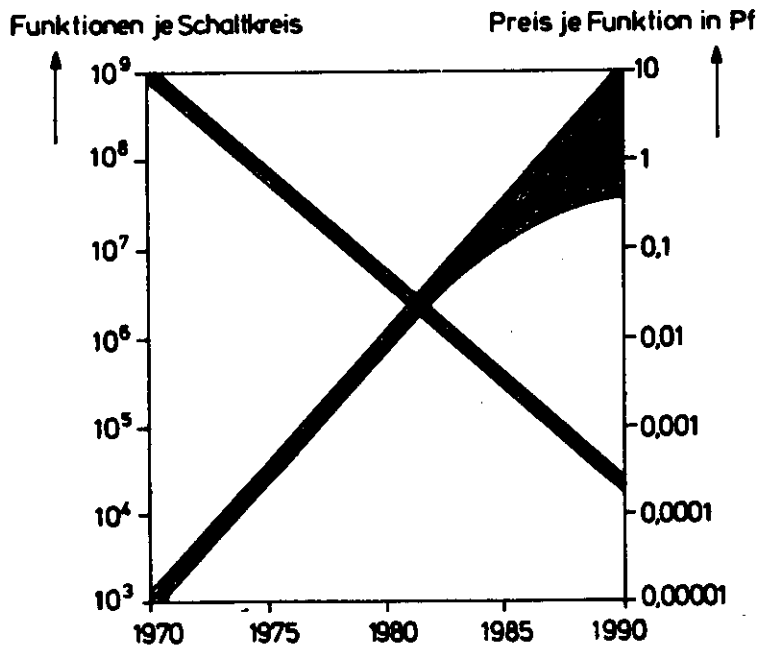


Die gleiche Struktur wie in Bild 34, jedoch nachchemisch gestzt. Es tritt ein starkes Unterätzen auf, außerdem entstehen Stege durch Blasenbildung, deutlich erkennbar bei der 2- $\mu\text{m}$ -Struktur (zweite von links)

26 Vorteile des Trockenätzens (hier Plasma-Ätzen)



28 Entwicklungsgeschichte der IC's



29 Zunahme der Integrationsdichte und Sinken der Kosten

ob 1990 30 Mill Funktionen oder eine Milliarde integrierbar sind, die Grenze wird nicht mehr die Technologie, sondern die Anwendung setzen. Es erscheint auch fraglich, ob die gezeigte Kostenentwicklung überhaupt genutzt werden kann, zumal sie nur erreichbar ist, wenn diese mikroelektronischen Schaltungen in großen Serien gefertigt werden. Der Anfangsaufwand für die Fertigung solch komplexer Schaltungen ist in der Entwicklung, der Fertigung und der Prüfung so hoch, daß er die Kosten besonders bei kleinen Mengen bestimmt. Der Aufwand beträgt heute bereits bis zu einigen Millionen DM und wird bei den sehr teureren Techniken der  $1\mu\text{m}$  Strukturen noch stark steigen. Daraus könnte man folgern, daß es für die höchstintegrierten Schaltungen gar keine Anwendungen mehr gibt, das ist jedoch nicht der Fall. Wir werden diese Schaltungen brauchen, wenn man technische Intelligenz mit großen Speichersystemen entwickelt und diese durch Redundanz funktionssicher machen muss.

Der Schaltungsentwurf solch riesiger Systeme mit etlichen Millionen Transistorfunktionen bereitet beträchtliche Schwierigkeiten. Jedoch fängt nicht jeder Entwurf beim Einzeltransistor an, die Hersteller haben inzwischen in jeder Technologie eine Bibliothek von Masken, die z.B. alle Sorten von FF's enthalten, aber auch zusammengesetzte Register, Decoder, Multiplexer, Adder usw., die in sich logisch geprüft sind, deren Timing genau bekannt ist und deren Chip Layout inklusive der Anschlußpins definiert sind. Beim Entwurf eines größeren Systems, der mit einer CAD(Computer Aided Design)-Anlage durchgeführt wird, können diese Teilchips, meist Makros genannt, aus der Bibliothek geholt, geeignet verbunden werden und so zu einem größeren Chip zusammengesetzt werden. Diese Methode wird meist für den ersten Entwurf benutzt, da aber oft nicht der ganze Inhalt der Makros für die Schaltung benötigt wird, sondern nur Teile, verbraucht man zu viel Chipfläche für das System. Deshalb wird nach dem Test des ersten Entwurfs ein Redesign durchgeführt mit dem Ziel, Chipfläche und damit Kosten zu sparen. Oftmals bringt auch der erste Hersteller den Originalentwurf als Chip auf dem Markt, um der erste zu sein, das Redesign wird dann von den second source-Herstellern durchgeführt.

Aber das Vorhandensein solcher Makrobibliotheken bringt natürlich die Hersteller auf die Idee, diese auch den Kunden anzubieten mit dem Ziel, ihn zu veranlassen, seine eigene Problemlösung als integrierte Schaltung zu versuchen. Die Technik heißt Gate Arrays, Schaltungen mit mehr als 1000 Gates auf einem Chip, die vom Kunden selbst logisch verknüpft werden können. Diese Lösung für kleine Serien hat den Vorteil, daß das Grundelement ein Standard Serien Produkt ist und die spezielle elektrische Schaltungstechnik durch die letzte Maske eingebracht werden kann. Die Makrozellen sind auf dem Chip in Reihen und Spalten angeordnet mit Zwischenräumen für Verdrahtungsbahnen. Die Zellen enthalten die logischen Makro-Schaltungen aus der Bibliothek, in unserem Beispiel 48 Zellen. Am äußeren Rand (Bild 30) befinden sich die Anschlußflecken (Pads) zur Befestigung der Anschlußdrähte. Dazwischen liegen Ein- bzw. Ausgangsschaltungen. Während der Herstellung werden sie auf die jeweils gewünschte Funktion festgelegt. Sie verbinden als Peripherieschaltungen das Array mit der Außenwelt, in unserem Beispiel über 68 Pins.

Aus dem in Serie vorfabrizierten Vorprodukt wird ein funktionstüchtiges Bauelement, wenn das Verdrahten der Schaltungselemente entsprechend dem vom Kunden gelieferten Schaltungsplan durch Aufbringen der Al-Leitbahnen geschieht. Für den Kunden stehen rechnergestützte Entwurfshilfen für die Designphase mit umfangreicher Software zur Logiksimulation zur Verfügung. Der Kunde wird geschult, um die Funktionsblöcke mit Komponenten aus der Makrobibliothek zusammensetzen. Dann kann er das Netzwerk in den Rechner geben und sowohl Logik als auch die Timingsimulation durchzuführen. Auch Testsignalfolgen legt er fest. All dies geschieht ohne Kenntnis der Integrationstechnologie. Der ganze Ablauf ist in Bild 31 zusammengefaßt. Wenn der Kunde alle diese Dinge durchgeführt hat, trägt er natürlich auch die Verantwortung für die richtige Funktion der Schaltung.

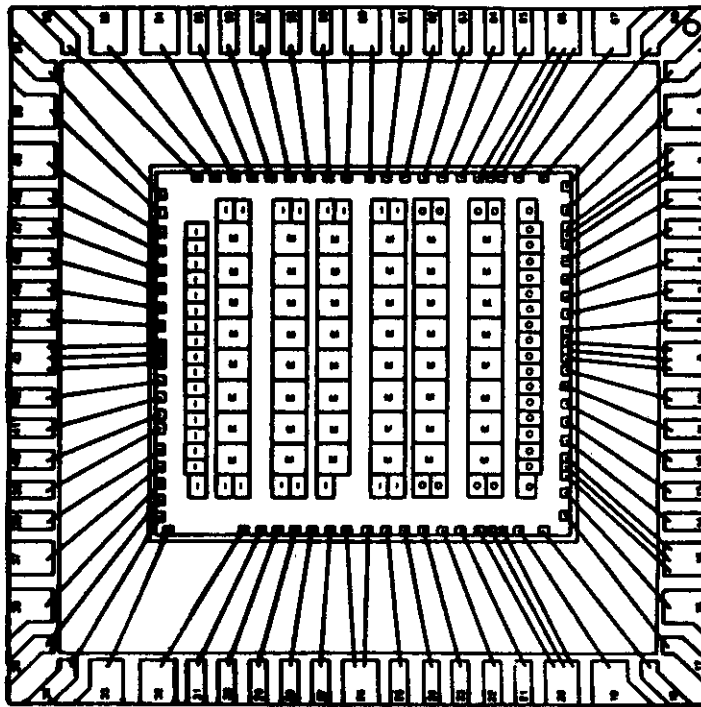
Die Hersteller bieten den Kunden diese Arrays in verschiedenen Technologien und verschiedenen Komplexitätsgraden an. Bis zum ersten Chip, den der Kunde in ca. 8 Wochen erhält, muß er zwischen 30 und 100 KDM hauptsächlich für CPU-Zeit auf den Firmenrechnern und Softwarekosten zahlen. Löhnen kann sich das Verfahren bei Stückzahlen zwischen 1000 und 30000, wobei die Chipkosten immer noch zwischen 200 und 10 DM liegen. Für höhere Stückzahlen sollte man spezifische Integration wählen, wobei die Chipfläche wesentlich kleiner werden kann (Preis, Ausschuß).

Der erste Rechner, der aus Gate-Arrays zusammengesetzt ist, ist die VAX 750 von DEC.

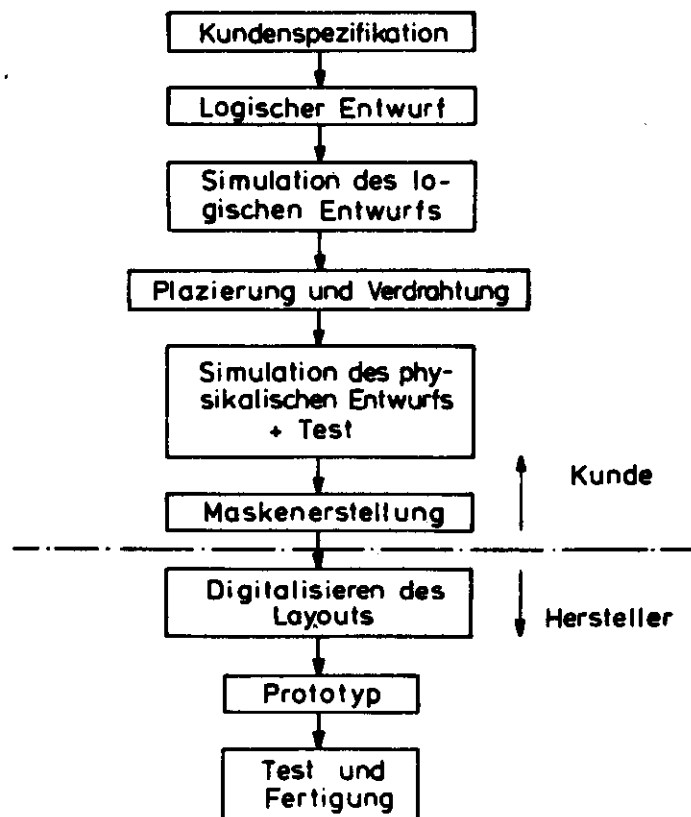
#### 4. Packaging.

##### 4.1 Montagetechnik.

Da die einzelnen Chips im Endstadium nur  $200\mu\text{m}$  dick sind, sind sie gegenüber mechanischen Belastungen ausserordentlich stör anfällig. Man baut die Chips normalerweise in Gehäuse ein, die eine leitende Verbindung



30 Makrozellen Anordnung



31 Entwurfsablauf für Gate-Arrays

zwischen der integrierten Schaltung und den Anschlüssen gewährleisten und nach Möglichkeit hermetisch verschlossen sein sollten.

Als Gehäusetypen kennt man Rund-, Flach- und DIL-Gehäuse (dual-in-line). Rund- oder auch TO-Gehäuse (TO:transistor outline) werden häufig bei Einzeltransistoren oder auch integrierten Schaltungen mit nicht zu vielen Anschlüssen eingesetzt (Bild 32). Flachgehäuse zeichnen sich durch eine geringe Bauhöhe aus. Sie bestehen aus Kunststoff oder Keramik und haben u.a. 10 ... 14 Anschlüsse (siehe Bild 33). DIL-Gehäuse werden am häufigsten für die Verkapselung von integrierten Schaltungen verwendet. Sie bestehen aus einem rechteckigen Gehäuse, deren mittlerer Pin-Abstand 1/10 Zoll (2,54 mm) ist. Man kennt Ausführungen mit 14, 16, 18, 24, 28, 40 oder 64 Anschluss stifen. Die Gehäuse können kunststoffummantelt sein oder aus Keramik (Cerdip) bzw. speziellen Glassorten bestehen (siehe Bild 34).

Die Siliziumchips müssen, bevor sie z.B. kunststoffvergossen werden, auf einem Träger befestigt werden. Dies dient einmal dazu, die Anschlussdrähte zwischen einem Chip und dem Gehäuse definiert anbringen zu können; aber auch als termischer Kontakt zur Wärmeableitung und in manchen Fällen auch als elektrischer Kontakt zwischen dem Substrat der integrierten Schaltung und einem Gehäuseanschluss.

Eine weitverbreitete Möglichkeit, den Chip auf einem Substrat zu befestigen, ist die Legierungstechnik. Hierbei wird ein Gold-Silizium-Eutektikum (3,6 Gew.-% Silizium und 96,4 Gew.-% Gold) auf die Rückseite des Chips aufgebracht. Bei Temperaturen von etwa 400° C schmilzt das Eutektikum. Ein weiteres, häufig verwendetes Verfahren ist die Klebtechnik, bei der mit Hilfe von Epoxidharzen eine fest haftende Verbindung zwischen der Unterlage und dem Chip geschaffen wird.

#### 4.2 Kontaktierung.

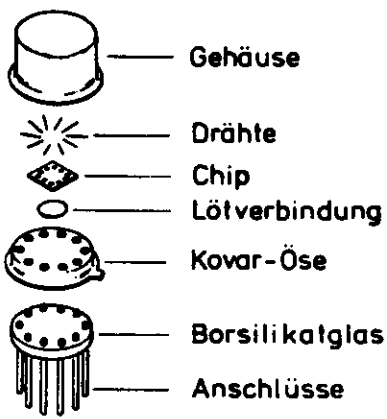
Um die Kontaktflecken auf dem Chip mit den Anschlüssen des Gehäuses zu verbinden, kennt man heute im wesentlichen zwei Verfahren: Thermokompression (vorwiegend für Golddrähte) und Ultraschall (vorwiegend für Aluminiumdrähte).

Das bekannteste Thermokompressionsverfahren ist die Nagelkopfkontaktierung (nailhead-bonding oder ball-bonding). Hierbei wird ein Golddraht durch eine Düse geführt, und das Drahtende wird mit Hilfe einer Knallgasflamme oder eine elektrischen Funkenstrecke zum Schmelzen gebracht, wodurch sich eine Kugel bildet, deren Durchmesser etwa das Doppelte des Drahtdurchmessers beträgt (Bild 35). Anschliessend bringt man das auf etwa 250° C erwärmte Goldkugelchen in Kontakt mit dem Kontaktfleck auf der integrierten Schaltung und presst die Düse auf das Kugelchen. Hierdurch kommt es zur festen Verbindung mit dem Kontaktplättchen. Nach dem Kontaktieren des Golddrahtes auf einer Kontaktstelle des Chips wird die Düse angehoben und zu dem Anschlußstift des Gehäuses geführt, wo eine Keilkontaktierung vorgenommen wird.

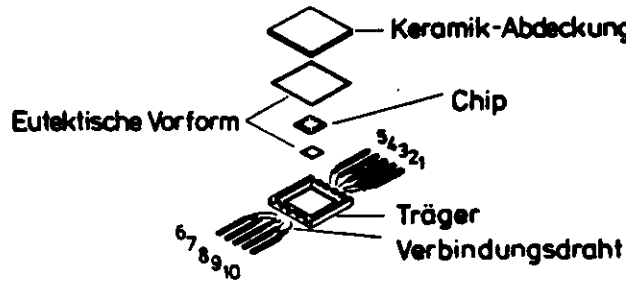
Beim Ultraschallverfahren wird bei Zimmertemperatur gearbeitet. Der Ultraschall dient einmal dazu, die Oxidschicht auf dem zu verbindenden Aluminiumdraht zu zerstören und durch Reibung des Drahtes mit der Kontaktstelle die notwendige Reibungswärme zu erzeugen, um dann unter Druckeinwirkung eine Keilkontaktierung vorzunehmen (Bild 36).

Da diese Verbindungstechniken nur halbautomatisch durchgeführt werden, bedeutet dies bei der Kalkulation der integrierten Schaltung einen erheblichen Kostenfaktor. Man hat daher bereits Überlegungen angestellt, wie sich die Kosten senken lassen. In einem Fall, der sogenannten Flip-Chip-Technik werden am Rande des Chips Kontaktflecken angebracht, die erhöht ausgeführt sind und mit Hilfe von Löttechniken mit dem Träger verbunden werden (Bild 37). Bei der Beam-Lead-Technik werden an die einzelnen Kontakte Stege angebracht (Bild 38), die später mit dem Träger kontaktiert werden.

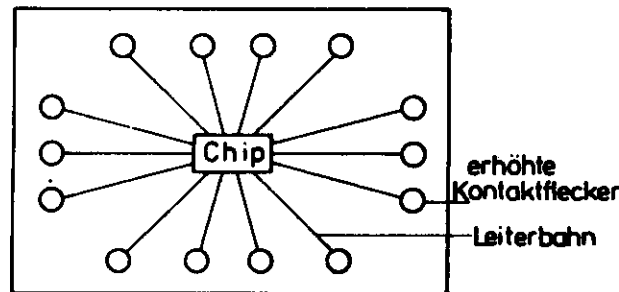
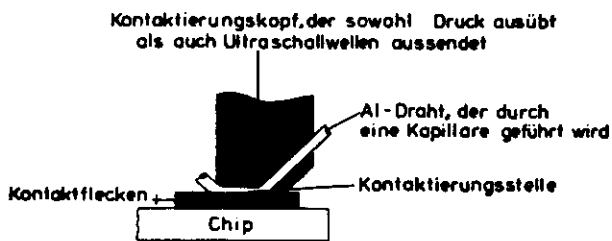
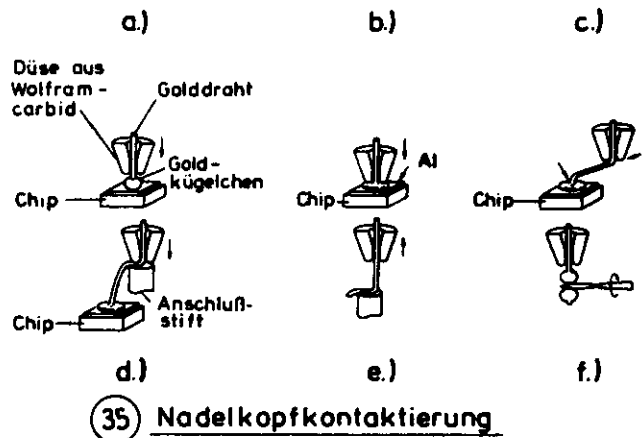
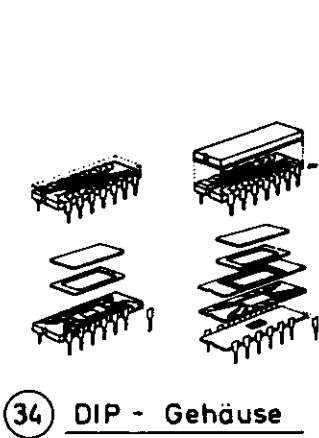
Die bisher erwähnten Gehäuse betrafen Chips, die kleine bis mittlere Integrationsgrade enthalten, wie sie bisher eingesetzt wurden. Die Großintegration zwingt jedoch dazu, Gehäuse mit wesentlich mehr Pins zu entwickeln. Ab etwa 1985 werden 200-300 Pins pro Gehäuse benötigt, die alten DIP-Gehäuse werden dann nur mit SSI- und MSI-Chip gefüllt sein. Diese DIP-Gehäuse haben Pinabstände von 100 mil, aus Platzgründen sind z.Zt. neue DIP's mit 50 mil teils in der Erprobung, teils in der Verwendung. Die Mini-DIP's, wie sie auch genannt werden, gibt es wie die normalen DIP's in 8, 14, 16, 20, 28, 40, 48 und 64 pin-Gehäusen, z.B. den Mikroprozessor Z8 von Zilog kann man so haben. Obgleich sie schon wesentlich weniger Fläche als die Normal DIP's beanspruchen, kann man weiteren Platz gewinnen, wenn man nicht nur an 2, sondern an allen 4 Seiten Kontakte anbringt. In dieser Bauform hat sich der Leadless Carrier Typ durchgesetzt, sowohl in Keramik als auch in Plastikform. Der Chip wird durch starke Federn in seiner Position gehalten, die Drahtanschlüsse dadurch auf die Chip-Anschlußbahnen gepresst und so der Kontakt hergestellt. Viele Firmen verkaufen den



32 Aufbau eines TO5-Gehäuses mit 10 Anschlüssen



33 Aufbau eines Keramikflächgehäuses



Chip so als Keramikfläche mit Anschlußbahnen. Wenn man den Anschlußabstand von 50 mil auf 20 oder 25 mil senkt, kann man in einem 1" x 1"-Gehäuse 150-180 Pins unterbringen. Für hohe Leistungen werden Kühltürme auf die Chips gesetzt (Bild 39). Um zu höheren Anschlußdichten zu kommen, muß man die Pins in mehreren Reihen staffeln, im Extremfall die ganze Fläche des montierten Chips mit den Anschlußbahnen dafür benutzen. Daher planen heute die meisten Firmen solche Pingitterarrays. Diese von IBM Ende der 70er Jahre vorgestellte Technik ist ein Multilayer Träger mit einer Pinmatrix im 100 mil-Abstand. Der Chip mit kleinen Löthügeln wird umgekehrt auf die Anschlußplatte gelötet. Diese Keramikplatte hat bis zu 23 Lagen und bis zu 261 Pins. (Bild 40) zeigt als Beispiel einen Chip, der die gesamte Logik einer IBM 370-CPU enthält mit 12,5 mm Kantenlänge auf einem Keramikträger von 36 mm Kantenlänge und 216 Anschlüssen. Der Träger ist dreilagig mit 0,01 mm Isolation zwischen den Lagen. Die Anschlußdrähte sind zwischen 1/40 und 1/3 mm dick. Verschiedene Techniken werden entwickelt, um die Chips mit dem Träger zu verbinden. Beim TAB (Tape Automated Bonding) (Bild 41) werden die Chips mit den metallischen Anschlußhöckern auf das innere Anschlußmuster in einem Arbeitsgang gebondet. Dabei können 1 oder mehrere (im Beispiel 4) Arrays auf eine Keramikträgerplatte montiert werden, z.B. 4 x 16 bit CPU's, die zu einer 64 bit CPU verbunden werden.

Auf diese Weise werden wir in den nächsten Jahren in der Lage sein, Chips mit bis zu 300 Anschlüssen einzubauen. Natürlich müssen auf den Karten, auf denen die IC's untereinander verbunden werden, die Bahnen entsprechend dünn sei, Feinstleiteteknik heißt dann 10-20  $\mu$ m dicke Bahnen.

### 5. Testen.

Ein entscheidender Gesichtspunkt der Größtintegration ist die Steigerung der Bauelementeverlässigkeit, denn die integrierten Verbindungstechniken der auf dem Chip vereinten Elemente ist wesentlich zuverlässiger als eine externe Verdrahtungstechnik. Das darf nicht darüber hinwegtäuschen, daß das Einzelement auf der Schaltung kaum noch erreichbar ist, man muß also versuchen, das hochkomplexe System auf sämtliche betrieblich denkbaren Möglichkeiten zu testen. Andererseits drängt sich sofort die Frage auf, ob es nicht möglich ist, Fehler auf dem Chip zu reparieren, z.B. dadurch, daß man zusätzliche, also redundante Elemente auf dem Chip unterbringt, die man nachträglich an Stelle der fehlerhaften einschaltet. Solche Verfahren werden schon bei Speichern benutzt.

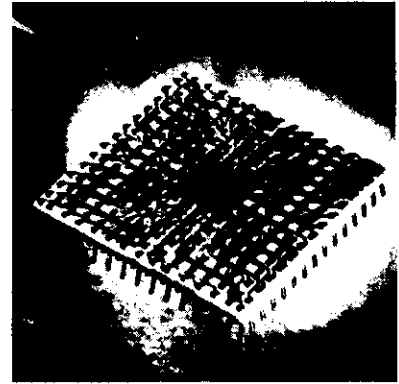
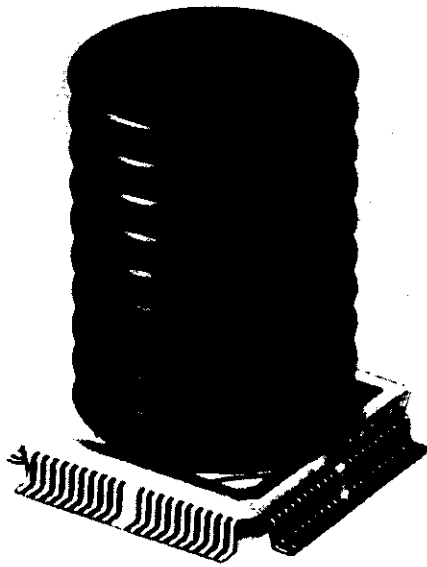
Beim Systemtest des gefertigten Chips hängt die Art und der Aufwand des Tests ganz von der Komplexität der Chips ab.

Automatische Testsysteme für MSI- bzw. LSI-Schaltkreise arbeiten heute mit programmierbaren Puls-Generatoren (Bild 42), die über Mikroprozessoren gespeicherte Zufallmuster, aber auch über gezielte Stimulation an einzelnen Pins einen programmierten Test erlauben. Die Zufallmuster mischen statistisch Bitmuster verschiedener Frequenzen, die als ganzes Muster periodisch wiederholbar sind. Diese sollen den Prüfling geeignet stimulieren und an allen Schaltelementen mindestens einen Schaltvorgang erzielen. Die Überwachung der Prüflingsreaktion erfolgt nach der Transition-Count-Methode, mit der die Pegelwechsel der Prüflingsausgänge gezählt und mit einem vorgegebenen Muster verglichen werden. Zuvor muß der Prüfling initialisiert werden, um geeignete Anfangsbedingungen zu erreichen.

Diese Testmethode verliert ihre Bedeutung mit hoher VLSI-Komplexität bei Schaltzeiten im Pikosekundenbereich, da über die Schaltmatrix unterschiedliche Laufzeiten zu den verschiedenen Pins das Ergebnis verfälschen. Neuere Testmethoden erzeugen daher (Bild 43) für jeden Pin getrennt ein programmierbares Testpattern, das über einen eigenen Pulsgenerator und eigenen Pegeltreiber nur auf den einen Pin geht. Dessen Reaktion wird dann gemessen und vom Rechner ausgewertet. Diese Geräte erzeugen Taktfrequenzen bis zu 80 MHz bei einer Genauigkeit von etwa Hundert Pikosekunden. Sie kosten allerdings inklusive umfangreicher Software 5 bis 10 Mill. DM.

### 6. Speicher.

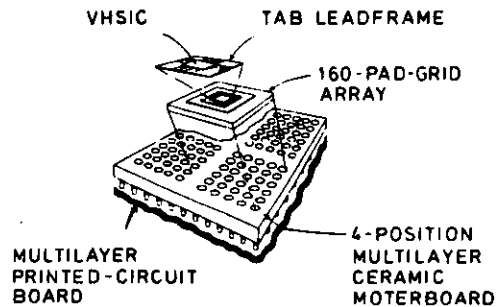
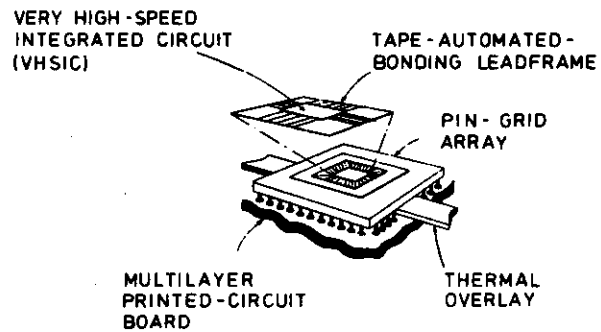
Damit man alle Vorteile der Großintegration ausnutzen kann, sind hohe Stückzahlen erstrebenswert. Millionenzahlen bei CPU's oder Peripheriebausteinen sind nicht so leicht zu erzielen, wohl aber bei Speichern, die von allen Rechnerherstellern verwendet werden können. Die großen Umsätze macht der, der als erster den nächstgrößeren Speicher auf den Markt bringt. Während in den vergangenen Jahren meist Speicher aus bipolaren Elementen gebaut wurden, hat sich das Geschäft fast vollständig auf MOS-Speicher verlagert. Die beiden Hauptgründe sind die hohe mögliche Integrationsdichte und der geringe Stromverbrauch ohne Signal, d.h. im Standbybetrieb (einige nW/bit gegenüber einige mW bei bipolarer Technik). Der Nachteil der MOS-Speicher ist ihre geringere Geschwindigkeit (Faktor(3-5) mal langsamer). Aber heute sind auch bei MOS-Speichern von 64K bits Zugriffszeiten von 100 ns üblich.



Tightly spaced. Amdahl's 580 computer uses this 84-lead ceramic chip-carrier with leads on 25- rather than 50-mil centers to house an emitter-coupled-logic large-scale integrated circuit. The tower-like heat sink dissipates up to 4 1/2 watts with air cooling.

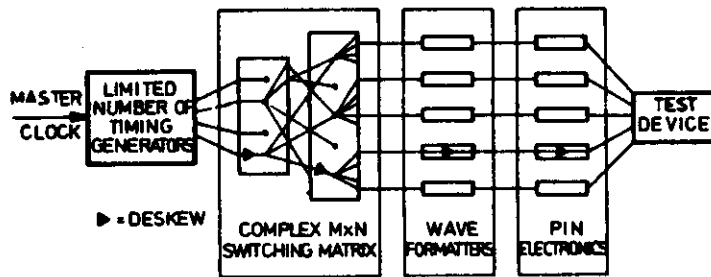
39 IC mit Kühlturm

40 Pin Grid Array

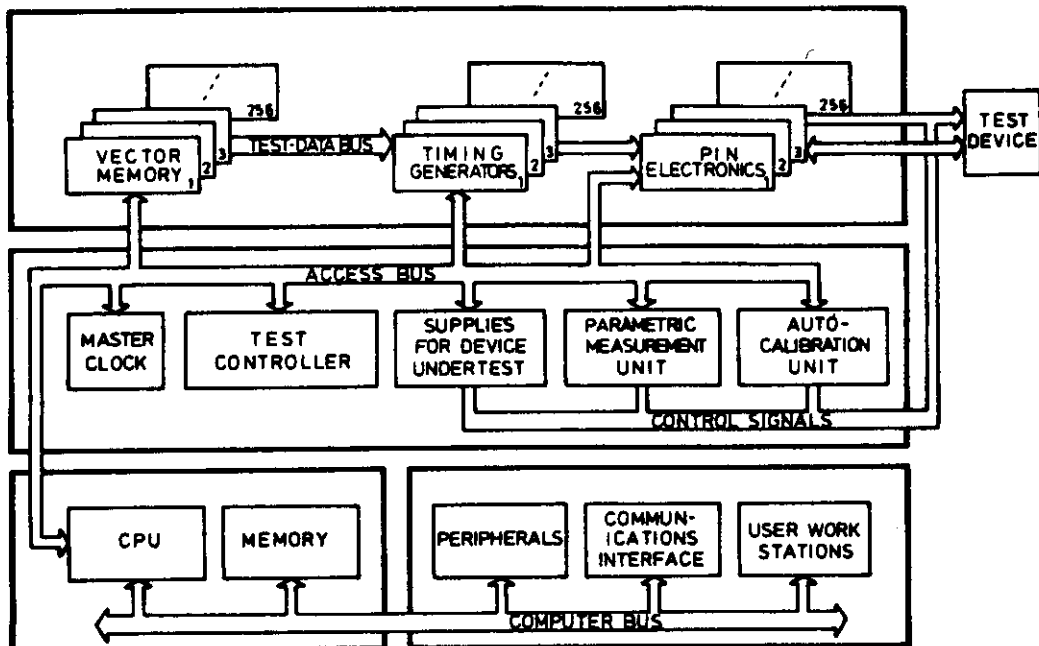


41 Kombinationen von Pin Grid Arrays





④2 Bisherige Prüfmethode der Chips



④3 Prüfmethode für VLSI-Chips

Der Umsatz an Speichern steigt rapide (Bild 44), er liegt jetzt bei einige Mrd \$ und wird in den nächsten 4 Jahren sich verdoppeln, z.Zt. gibt es 256Kbit-Chips, der 1 M bit-Chip wird voraussichtlich im nächsten Jahr als Muster vorgestellt, die Produktion wird 86/87 starten.

Aus Bild 44b unten erkennt man das Fallen der Preise in mCent/bit bei steigenden Umsatzzahlen, schon 1985 werden je nach Typ 3 bis  $25 \times 10^{-3}$  Cent/bit erwartet, das entspricht einem DRAM Preis von 16 \$ bei 256K bit.

Der Aufbau der vier Speichertypen und ihr Funktionsablauf ist leicht zu verstehen (Bild 45). RAM's sind Speicher mit wahlfreiem Zugriff, ROM's/PROM's nur Lesespeicher. In ROM's und PROM's werden in zunehmendem Maße Softwarepakete untergebracht und zwar Teile des Betriebssystems, Compiler oder Interpreter sowie Anwendungssoftware.

### 6.1 Statische MOS-Schreib-Lese-Speicher.

Bild 46a zeigt den Aufbau einer MOS-Speicherzelle für den statischen Betrieb in n-Kanal-Technik. Für ein wortorganisiertes RAM sind nur sechs MOS-Transistoren erforderlich, wobei die beiden Transistoren T3 und T4 als Arbeitswiderstände geschaltet sind.

Bei statischen MOS-Speicherzellen besteht die Zelle aus zwei rückgekoppelten Invertern, die zwei stabile Zustände aufweisen. Dem einen Zustand ist ein 0-Signal, dem anderen ein 1-Signal zugewiesen.

Ist bei der Schaltung der Transistor T1 leitend, ist der T2 gesperrt. Durch den Gateanschluss mit dem jeweils gegenüberliegenden Drainanschluss ergibt sich der stabile Zustand, der nur durch ein entgegengesetztes Signal an der Schreib-Lese-Leitung geändert werden kann.

Die beiden Transistoren T5 und T6 dienen zur Ansteuerung der Speicherzelle. Liegt auf der Zeilenleitung ein L-Pegel, sind diese Transistoren gesperrt. Der Zustand der Speicherzelle kann sich nicht ändern, da die Zelle als isoliert betrachtet werden muß.

Bei einer Adressierung liegt auf der Zeilenleitung ein H-Pegel. Zwischen Zelle und den beiden Schreib-Lese-Leitungen besteht eine elektrische Verbindung. Nun kann man entweder eine Information einschreiben oder sie zerstörungsfrei auslesen. Danach isoliert die Zeilenleitung wieder die Speicherzelle. Ist ein 1-Signal gespeichert, kann nur ein 1-Signal auf der 0-Schreib-Lese-Leitung den Speicherzustand ändern.

### 6.2 Dynamische MOS-Schreib-Lese-Speicher.

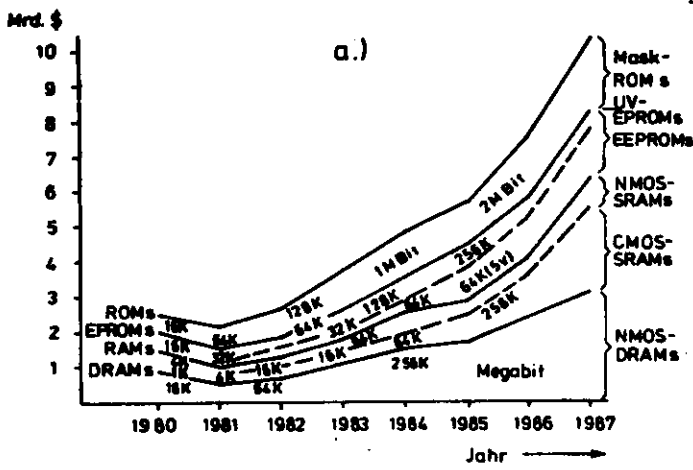
In der Massenspeicherung von Daten setzt man immer mehr RAMs in dynamischer MOS-Technik ein. Der Grund ist die hohe Integration, da jede Speicherzelle nur noch aus einem Transistor besteht. Bild 46b zeigt den Aufbau einer 1-Transistor-MOS-Speicherzelle mit einem kleinen Kondensator zwischen Transistor und Masse (Substrat). In diesem Kondensator ist die Information als Kondensatorladung gespeichert. Durch die Zeilenleitung wird die Speicherzelle angewählt. Ist in dem Kondensator ein 1-Signal abgespeichert worden, wird eine Spannung auf die Datenleitung geschaltet. Der Leseverstärker erkennt einen H-Pegel und schaltet den Ausgang entsprechend auf 1-Signal. Ist keine Ladung in dem Kondensator vorhanden, erkennt der Leseverstärker ein 0-Signal. Beim Einschreiben einer Information wird mit einem 1-Signal eine Ladung in dem Kondensator auf- oder bei einem 0-Signal abgebaut. Dies ist eine Ladung oder Entladung durch den Schreibverstärker.

Der Nachteil bei den dynamischen RAM's ist die kontinuierliche Nachladung der Speicherzellen durch einen Refresh-Zyklus. Wegen der unvermeidbaren Leckströme muß die Speicherinformation im Kondensator periodisch regeneriert (refresh) werden. Dieser Refresh erfolgt im Baustein durch einen Lesesyklus auf den Refreshadressen, wobei durch eine Ansteuerung eines Speicherelements immer eine gesamte Zeile regeneriert wird.

### 6.3 Bipolare Festwertpeicher.

Bei den Festwertspeichern muß man zwischen den maskenprogrammierten und den programmierbaren Speichern unterscheiden. Die ROM's (Read Only Memory) und die PROM's (Programmable ROM's) sind Nur-Lesespeicher, die keinen Schreibbetrieb kennen.

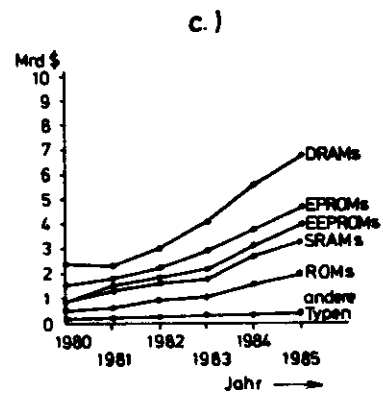
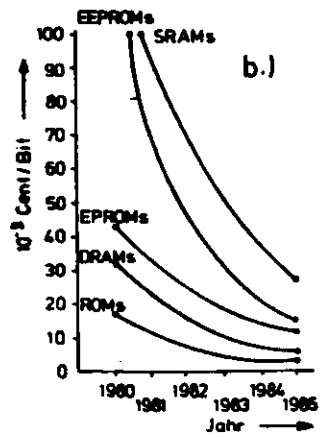
Bild 47 zeigt den Aufbau eines bipolaren Masken-ROM's. Im unprogrammierten Zustand bestehen zwischen den X- und Y-Achsen keine elektrischen Verbindungen. Wird eine Speicherzelle durch die Herstellung programmiert, besteht zwischen der Anode der Diode und der Y-Leitung eine elektrische Verbindung. Die Diode entkoppelt die Leitungen untereinander.



45

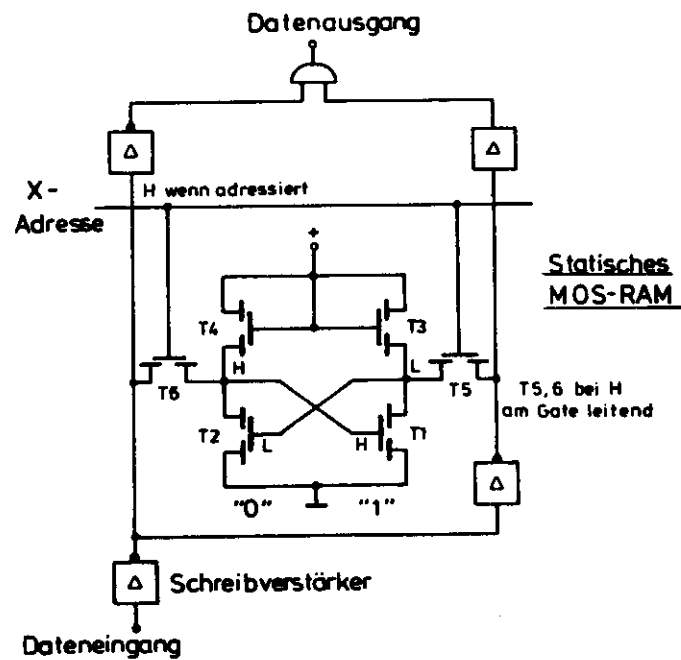
a.) Speichertyp	Bits/Chip	
	1983	1987
SRAM's	64 K	256 K
DRAM's	256 K	1 M
ROM's/PROM's	256K	2 - 4 M
EEPROM's	32K	64 - 256 K

44 Erwarteter Umsatz und Preisentwicklung bei Speichern



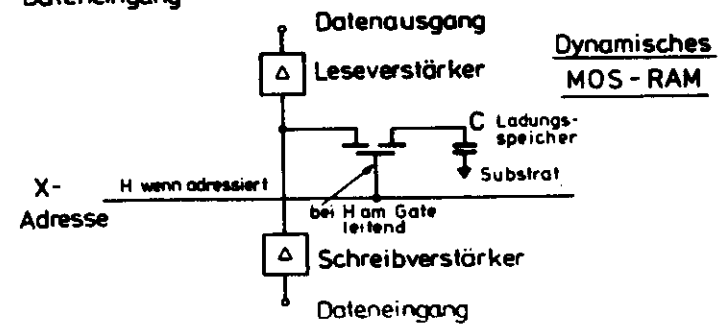
b.) Software in ROM's:

- Busprotokoll
- Busarbitrater
- Interrupthändler
- Bildschirmorientierter Monitor
- Pascalcompiler 32 - 40K Byte
- Basisinterpreter 20 - 32K Byte
- Floatingpoint handler
- Graphic



a.)

46



b.)

Die Leseverstärker bei einem Festwertspeicher haben eine H-Pegel an ihren Eingängen. Ist eine Adressierung durch einen L-Pegel auf einer Zeile erfolgt, schaltet der Eingang des Leseverstärkers auf L-Pegel, wenn eine leitende Verbindung besteht. Andernfalls bleibt der H-Pegel auf der Y-Leitung.

Im unprogrammierten Zustand sind nur die Dioden in der Speichermatrix vorhanden, die aber nicht abgeschlossen sind.

Setzt man in die Zwischenräume Sicherungselemente ein, kommt man zu den programmierbaren Festwertspeichern nach dem Fusable-Link-Verfahren. Bild 48 zeigt den Aufbau einer programmierbaren Speichermatrix.

Im unprogrammierten Zustand sind die Sicherungselemente unbeschädigt, d.h. zwischen den X- und Y-Achsen sind intakte Sicherungsbrücken vorhanden. Diese Brücken sind kleine und dünne Drähte, also Sicherungsdrähte (fusible link). Soll eine Speicherzelle programmiert werden, muß der Sicherungsdraht durch einen Stromimpuls abgeschmolzen werden.

Die Programmierung dieser Sicherungselemente ist nicht nur einfach, sondern dauert nur wenige Millisekunden.

Man legt auf die  $Y_0$ -Achse eine positive Betriebsspannung und schließt die  $X_1$ -Achse an Masse. Es fließt ein Strom und das Sicherungselement wird abgeschmolzen. Nach dem Abschmelzen der Brücke ist keine elektrische Verbindung mehr zwischen den beiden Achsen vorhanden.

Das Sicherungselement ist ein etwa  $0,02 \text{ mm}^2$  dicker Nickel-Chrom-Widerstand, der durch den Stromfluß abgeschmolzen wird. Die Programmierung verläuft hier anders als bei den maskenprogrammierten Festwertspeichern. Der Anwender selbst programmiert sich den Baustein nach seinen Anforderungen.

#### 6.4 Programmierbare MOS-Speicher.

Bei den programmierbaren MOS-Speichern muß man zwischen den maskenprogrammierbaren Festwertspeichern, den EPROM's (Erasable PROM's) and den EAROM's (Electrically Alterable ROM's) unterscheiden.

Bild 49 zeigt den Aufbau eines maskenprogrammierten Festwertspeichers in MOS-Technik. Für die Programmierung liefert der Kunde einen Lochstreifen mit seinem Programm. Ein Plotter zeichnet von diesem die Maske für die Herstellung.

Bild 50a zeigt den Aufbau einer Speicherzelle für EPROM's, die man als FAMOS-Speicherzelle (floating-gate-avalanche-injection- MOS) bezeichnet. Die Speicherzelle besteht aus zwei Gattern. Das erste ist das floating-gate, ein bewegliches und das andere ist das Ansteuerungsgatter für die Speicherzelle.

Das floating-gate speichert die Ladung; die Information. Während des Programmierens werden durch Injektion energiereicher Elektronen die Speicherzelle programmiert. Die Elektronen können nicht abfließen, da das bewegliche Gate isoliert zwischen den Halbleiterschichten liegt. Mittels des Avalanche-Effektes (Lawine), der durch eine hohe Programmiervoltage von  $U_p = 26 \text{ V}$  entsteht, tritt in der FAMOS-Zelle dieser Injektionseffekt auf.

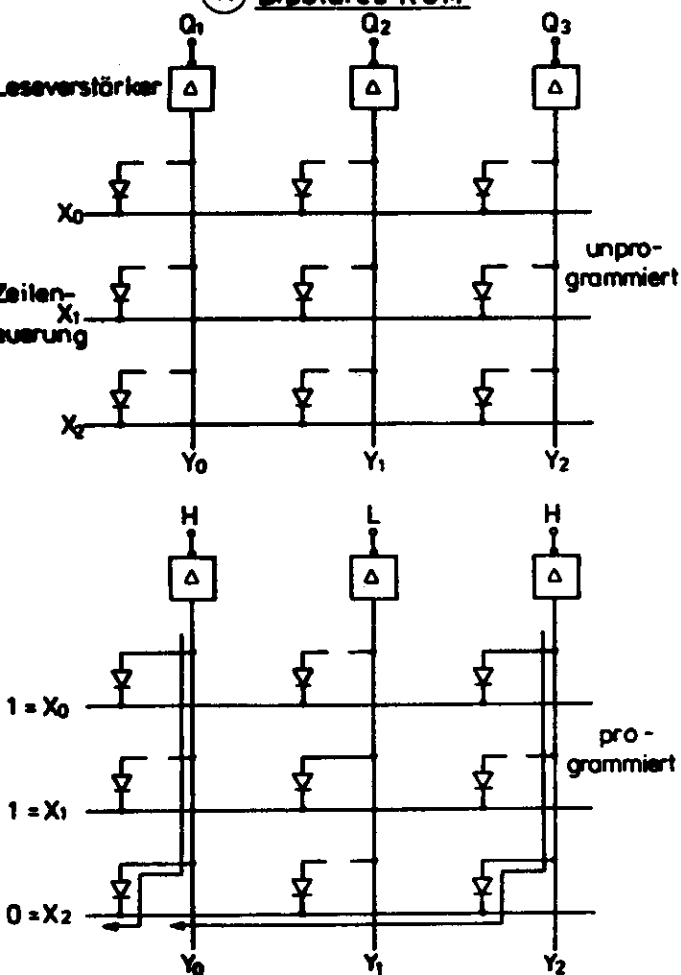
Bei einem UV-Licht von 250 nm-Wellenlänge lösen Photonen die induzierte Ladung auf. Die Löszeit liegt zwischen fünf und zehn Minuten. Während bei den EPROM's immer nur der gesamte Speicherinhalt zu löschen ist, kann man bei den EAROM's entweder nur ein Bit oder eine Wortzeile ändern.

Den Aufbau solcher Speicherzelle zeigt Bild 50b. Die Speicherzelle besteht aus einem MNOS-Transistor (Metal-Nitride-Oxide-Semiconductor). Zwischen dem Gateanschluß und der  $\text{SiO}_2$ -Schicht liegt das bewegliche Gatter zur Speicherung der Information. Dabei handelt es sich um eine Zweilag-Isolation, die aus Siliziumnitrid und Siliziumoxid besteht.

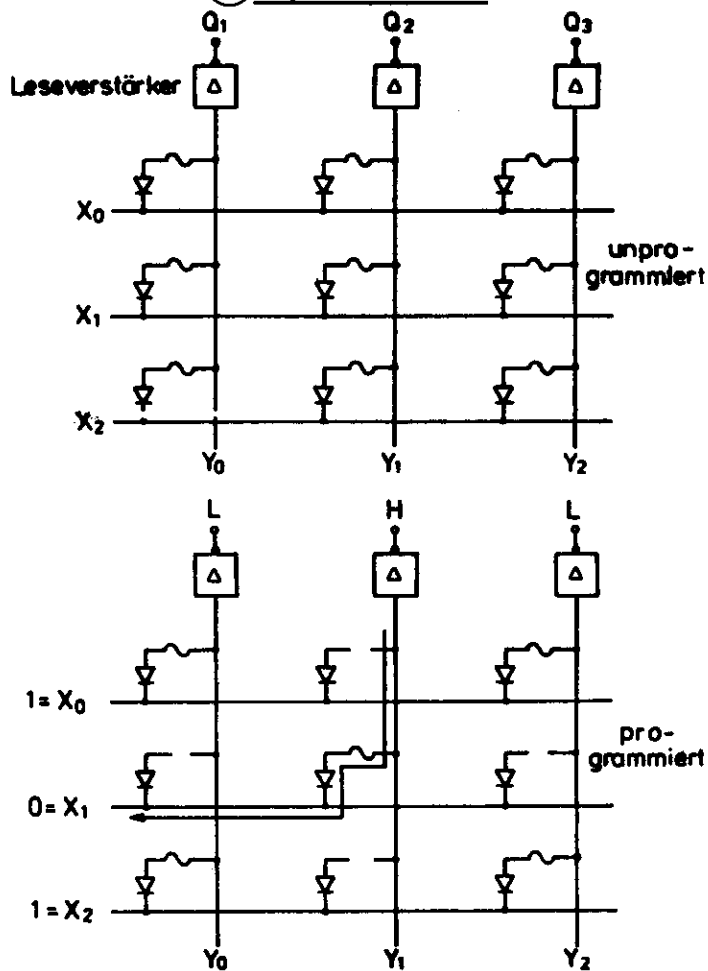
Durch Anlegen einer positiven oder negativen Programmiervoltage läßt sich das Gate nach oben oder unten verschieben. Dadurch ändert sich gleichzeitig auch die Wirkung der Gatespannung auf das n-Substrat. Bei einer positiven Spannung am Gate werden Elektronen aus dem n-Substrat in das Siliziumnitrid gedrängt und daher ändert sich auch die Schwellenspannung des MNOS-Transistors. Bei einer negativen Spannung werden die Elektronen aus dem Siliziumnitrid in das n-Substrat zurückgedrängt. Durch die Verschiebung der Elektronen ergibt sich entweder ein Kanal zwischen den beiden p-Zonen (Source und Drain) oder der Kanal wird abgebaut.

Legt man das Gate des MNOS-Transistors auf die Masse, kann der MNOS-Transistor die Informationen speichern, selbst wenn man die Betriebsspannung abschaltet. Die Speicherung bleibt über lange Zeit erhalten, ist aber sehr temperaturempfindlich.

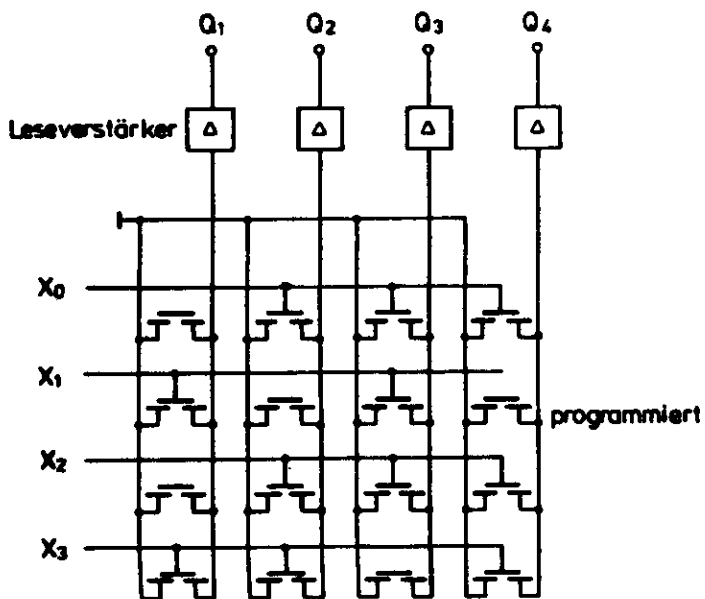
(47) Bipolares ROM



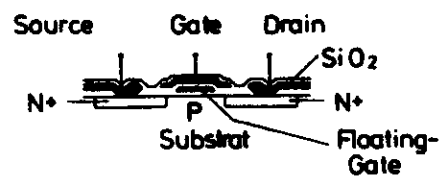
(48) Bipolares PROM



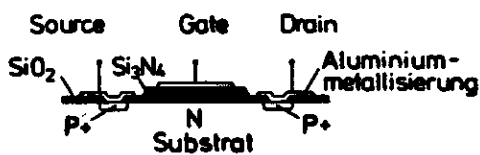
(49) MOS-ROM



a.) Speicherzelle für EPROMs



(50)



b.) Speicherzelle für EEPROMs

Das Löschen einer gespeicherten Information erfolgt durch Anlegen eines positiven Impulses, dessen Amplitude und Breite genau definiert sein muß. Durch den Löschvorgang wird die Schwellspannung auf 1-Signal gebracht.

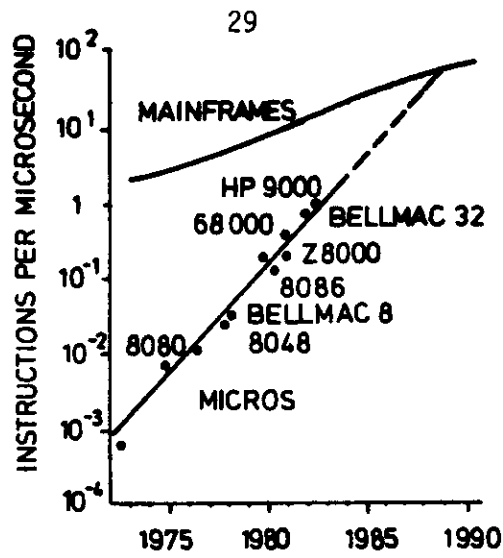
### 7. Zusammenfassung.

Die Halbleiterindustrie ist in der Lage, praktisch alle Teile des Rechnersystems in Form von Chips herzustellen. Dabei wird die Technologie der Herstellung so weit verbessert, daß mit hoher Dichte viele Millionen Transistorfunktionen auf dem Chip integriert werden können. Lasergesteuerte Versuchsanlagen sollen im nächsten Jahr schon 1000 Transistorfunktionen in 1s auf den Chip malen können. Das bedeutet, daß in einer Nacht sämtliche Schaltkreise des heute größten IBM-Rechners auf einer Si-Scheibe von 12,5 cm Durchmesser untergebracht werden können. Der heutige Supercomputer, der 25 Mill. DM kostet, könnte im Maßstab 1000:1 auf einer Kristallscheibe zu haben sein. Man erkennt, welche ungeheuren Möglichkeiten in solcher Technik stecken.

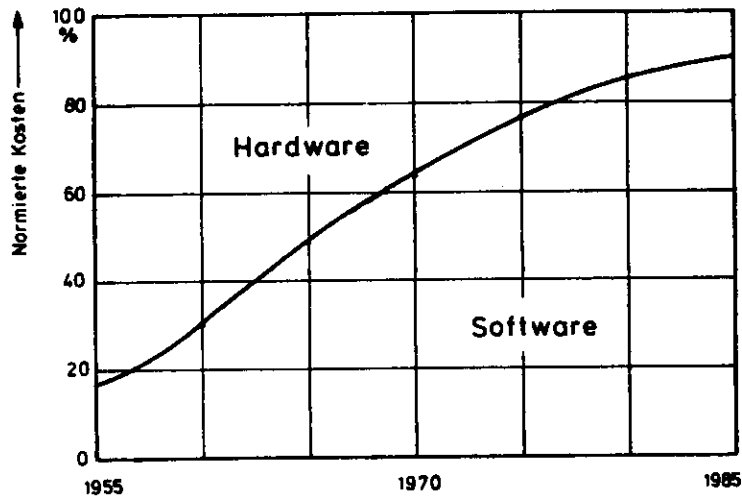
Auf der anderen Seite nähert sich die Leistungsfähigkeit heutiger Mikroprozessoren schon deutlich derjenigen der Großrechner (Bild 51). In den 90er Jahren werden diese Unterschiede schon verwischt oder verschwunden sein. Dabei wird die Hardware immer billiger, die Software als Arbeitsleistung immer teurer (Bild 52).

Dies ist nur möglich dank immer größer werdender Stückzahlen, besonders für Speichersysteme. Bis 1980 wurden weltweit nur etwa  $10^{13}$  Transistorfunktionen eingesetzt, kaum mehr, als der Kapazität von etwa 10 menschlichen Gehirnen entspricht; bis 1985 werden es etwa  $10^{14}$  sein, falls der Trend anhält, alle 5 Jahre 1 Größenordnung mehr. Um die Ausbeute groß zu halten, wird die Chipfläche verkleinert werden müssen, mehrere Hersteller denken an dreidimensionale Chips.

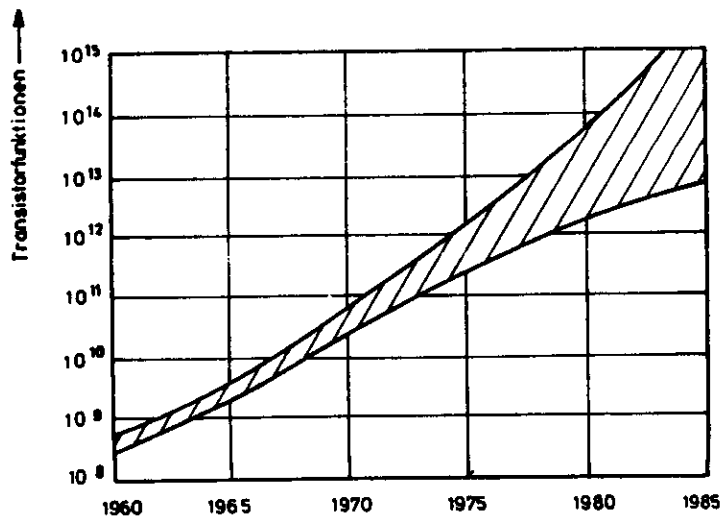
Um zu wesentlich höheren Geschwindigkeiten zu kommen, wird man neue Halbleitermaterialien benutzen, die z.T. heute schon im Test sind: GaAs mit wesentlich höheren Ladungsträgerbeweglichkeiten als beim Si, sowie supraleitende Josephson-Tunneleffekt Schalter, die im flüssigen He arbeiten und mit denen in IBM-Forschungsinstituten Schaltzeiten von etwa 20 Picosekunden erreicht wurden. Hiermit werden dann Ende unseres Jahrhunderts Rechner möglich sein, die einige 100 MIPs bis hin zu 1 GIPs verarbeiten mit Speichersystemen von einigen GBytes.



51) Rechenleistung der Mikros im Vergleich zu Mainframes



52) Anteile der Hard- und Softwarekosten bei Rechnern



53) Weltweit installierte Transistorfunktionen

