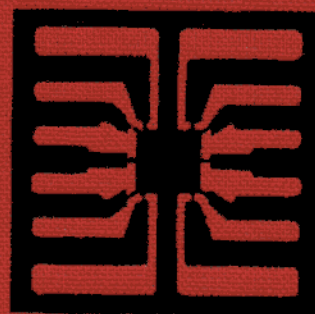




Rolf Enderlein

Mikroelektronik

*Eine allgemeinverständliche Einführung
in die Welt der Mikrochips, ihre Funktion,
Herstellung und Anwendung*

Mit 63 Abbildungen



VEB Deutscher Verlag der Wissenschaften · Berlin 1986

Verlagslektor: Gisela Zielas
Typographie: Gisela Deutsch
Umschlaggestaltung: Hans Wunderlich
© 1985 VEB Deutscher Verlag
der Wissenschaften, DDR-1080 Berlin,
Postfach 1216
Lizenz-Nr. 206 · 435/73/85
Printed in the German Democratic Republic
Gesamtherstellung:
VEB Druckerei «Thomas Müntzer»,
5820 Bad Langensalza
LSV 1159
Bestellnummer: 571 202 8
02500

Vorwort

Die Mikroelektronik gehört zu den bedeutendsten wissenschaftlich-technischen Errungenschaften unserer Zeit — viele halten sie für die wichtigste. Sie ist vergleichbar nur mit solch umwälzenden technischen Neuerungen früherer Epochen wie der Erfindung des Buchdrucks im 16. Jahrhundert, der Einführung der Dampfmaschine im 18. Jahrhundert und der technischen Erschließung der Elektrizität im 19. Jahrhundert. Wenn heute von wissenschaftlich-technischer Revolution gesprochen wird, so geht es dabei in erster Linie um die Mikroelektronik. Wie keine andere technische Neuerung unserer Tage wirkt sie in alle Bereiche des Lebens hinein und ermöglicht Dinge, die vorher unvorstellbar waren. Taschenrechner, Miniradios, elektronische Steuerungen in Haushaltsgeräten, Quarzuhren, elektronische Spielgeräte und programmierbare elektronische Rechner hoher Leistungsfähigkeit gehören zu ihnen. Aber die Mikroelektronik kann weit mehr. Ihre universelle Bedeutung beruht darauf, daß sie für etwas zuständig ist, was in seinem Stellenwert für das Zusammenleben der Menschen in der Gesellschaft nur noch mit der Energie und den stofflichen Materialien verglichen werden kann. Dieses «Etwas» ist die *Information*, und zwar im breitesten Sinne, in dem man sich diesen

Begriff denken kann. Informationsprozesse finden überall und ständig statt — in der Industrie, der Verwaltung, der Wissenschaft oder wo sonst zweckgerichtete materielle und geistige Vorgänge ablaufen. Information muß dabei gewonnen, verarbeitet, übertragen und gespeichert werden. Für die rationelle, schnelle und zuverlässige Durchführung dieser Informationsprozesse besteht ein starkes praktisches Bedürfnis. Die Mikroelektronik ist das Instrument zu seiner Befriedigung. Darin besteht ihre große Bedeutung. Solche oder ähnliche Worte über die Rolle der Mikroelektronik sind nicht gerade neu, und obwohl sich ihr Sinn nur dem fachlich Vorgebildeten voll erschließt, sind sie allgemein akzeptiert. Die Tatsachen sprechen für ihre Richtigkeit. Nicht jeder wird sich aber allein mit den Tatsachen, d. h. mit den für jedermann sichtbaren Auswirkungen der Mikroelektronik, zufrieden geben wollen. Er wird verstehen wollen, wie diese Wirkungen zustande kommen. Als Albert Einstein in einer Rede anlässlich der Eröffnung der 7. Funkausstellung 1930 in Berlin — bezogen auf das damals noch neue Radio — u. a. sagte: «Sollen sich auch alle schämen, die gedankenlos sich der Wunder der Wissenschaft und Technik bedienen und nicht mehr davon geistig erfaßt

haben als die Kuh von der Botanik der Pflanzen, die sie mit Wohlbehagen frißt»,

so war das sicher eine scherzhafte Übertreibung. Aber eine Wahrheit lag doch darin. Sie trifft heute auch auf die Kenntnisse über Mikroelektronik zu.

Die Mikroelektronik ist ein außerordentlich komplexes Gebiet. Man kann sich mit ihr unter einer Reihe verschiedener Aspekte beschäftigen. Von ihrer grundsätzlichen Funktion für die Gesellschaft war allgemein schon die Rede. Aber welche praktischen Bedürfnisse sind es im einzelnen, die die Mikroelektronik befriedigt und die zu ihrer Herausbildung geführt haben? Welche mikroelektronischen Erzeugnisse stehen heute zur Verfügung, und wie lassen sich diese in der Industrie, der Verwaltung, der Kommunikationssphäre, der Wissenschaft oder der Freizeit der Menschen anwenden? Wie werden diese Möglichkeiten bereits heute genutzt, und welche Entwicklungen sind für die Zukunft absehbar? Hierher gehören auch die sozialen Auswirkungen der Mikroelektronik, die — wie bei jeder technischen Neuerung, die in alle Sphären des gesellschaftlichen und individuellen Lebens hineinwirkt — außerordentlich vielseitig und tiefgreifend sind. Neben den Fragen der gesellschaftlichen Be-

dingtheit und Funktion der Mikroelektronik sind vor allem die Probleme ihrer technischen Wirkungsweise von Interesse. Wie funktioniert die Mikroelektronik, wie arbeiten die Halbleiterbauelemente, wie die aus ihnen aufgebauten Schaltkreise? Die Frage nach dem Funktionieren ist einerseits im Hinblick darauf zu stellen, welche physikalischen Prozesse in den Bauelementen, Schaltkreisen, Geräten oder Anlagen vor sich gehen, andererseits aber auch im Hinblick darauf, welche logischen Abläufe in diesen Systemen stattfinden und welche Vorschriften oder Programme aufgestellt werden müssen, um diese Abläufe in der für den jeweiligen Anwendungszweck erforderlichen Weise zu steuern. Es ist üblich, die elektronischen Schaltkreise, Geräte und Anlagen in ihrer *materiellen* Gestalt als *Hardware* und die für ihren Betrieb erforderlichen *geistigen* Voraussetzungen in Gestalt von Regelungen und Programmen als *Software* zu bezeichnen. Zu jedem mikroelektronischen System gehören gleichermaßen Hardware und Software. Man muß sich mit beiden beschäftigen, wenn man das System verstehen will. Wer sich über die Mikroelektronik informieren möchte, wird sich schließlich auch für die Frage interessieren, wie die mikroelektronischen Schaltkreise hergestellt werden. Dieser Aspekt ist für die Mikroelektronik sogar besonders wichtig. Mikroelektronik war und ist auch heute noch wesentlich die Kunst, Schaltkreise mit einer großen Anzahl von Bauelementen industriemäßig herzustellen. Die Zusammenfassung von 100000 Transistoren zu einem einzigen Schaltkreis, wozu heute die Mikroelektronik durchaus

in der Lage ist, war seit langem prinzipiell denkbar. Wenn dieser Schritt lange Zeit praktisch nicht vollzogen werden konnte, so lag das nicht daran, daß bestimmte physikalische Wirkprinzipien erst noch entdeckt werden mußten, sondern ausschließlich daran, daß die Kunst es zu tun, die Technologie, noch nicht beherrscht wurde. Der Durchbruch gelang 1960 mit der Herstellung der ersten *integrierten Schaltkreise* unter Anwendung der *Silizium-Planartechnologie*. Damit wurde die Entwicklung der Mikroelektronik eingeleitet. Auch die Zukunft der Mikroelektronik ist vorerst noch weitgehend eine Frage der weiteren Entwicklung ihrer Herstellungstechnologie.

Unter diesen und noch anderen Aspekten kann man sich mit der Mikroelektronik beschäftigen. Von ihnen ist in den zahlreichen Publikationen die Rede, die in den letzten Jahren zur Mikroelektronik erschienen sind. Sofern sich diese Publikationen aber überhaupt an den Nichtspezialisten wenden, steht in der Regel *ein* Aspekt, etwa die Anwendung oder die Herstellung mikroelektronischer Erzeugnisse, im Mittelpunkt. Eine solche Einschränkung geschieht aus guten Gründen und hat ihre volle Berechtigung. In diesem Buch wollen wir uns jedoch bewußt nicht auf *einen* der genannten Aspekte beschränken. Es soll versucht werden, zu den meisten das für ein allgemeines Verständnis Wichtigste zu sagen. Das Buch stellt unseres Wissens den ersten Versuch dar, das Gesamtgebiet der Mikroelektronik in allgemeinverständlicher Form zu behandeln. Es wendet sich an Leser, die, ohne über Spezial-

wissen aus einem Hochschulstudium zu verfügen, die Mikroelektronik begreifen wollen. Das Buch ist nicht für Spezialisten gedacht und soll kein Fachlehrbuch der Mikroelektronik sein. Solche Lehrbücher gibt es bereits in größerer Zahl. Die meisten sind für ihren Verwendungszweck, die Aus- und Weiterbildung von Ingenieuren und Naturwissenschaftlern auf dem Gebiet der Mikroelektronik, gut geeignet. Es war nicht die Absicht des Autors, die Zahl dieser Lehrbücher zu vergrößern. Das Buch soll das leisten, was Fachbücher nicht leisten können — dem nicht speziell vorgebildeten Leser eine Vorstellung von der Mikroelektronik zu vermitteln. Daran orientieren sich die Auswahl des Stoffes und die Form der Darstellung in erster Linie.

Im Mittelpunkt des Buches steht die Mikroelektronik, wie sie heute ist. Zwar werden auch die Triebkräfte dargestellt, die zu ihrer Herausbildung geführt haben, aber das geschieht in einem umfassenderen Sinne als durch das Nachzeichnen der technischen Entwicklungslinien. Auf letztere wird nur am Rande eingegangen. Wir konzentrieren uns auf die heute von der Industrie am breitesten genutzten Prinzipien, technologischen Verfahren und Anwendungen der Mikroelektronik. Dabei führen wir den Leser bis an den international neuesten Stand heran, verlagern aber das Gewicht nicht einseitig auf das erst in jüngster Zeit Erreichte. Ohnehin ist bei dem heutigen Entwicklungstempo der jeweils neueste Stand schnell veraltet. Was aktuell bleibt, sind vor allem die Grundlagen. Wer sie beherrscht, kann sich das Neue leicht aneignen.

Die Grundlagen werden deshalb in diesem Buch besonders betont. Im ersten Kapitel wird in allgemeiner Form dargestellt, was Elektronik ist und in welcher Beziehung sie zur Information und ihrer Verarbeitung steht. Das zweite Kapitel handelt von der Bedeutung der Mikroelektronik für die Gesellschaft. In Kapitel 3 beschäftigen wir uns mit den physikalischen Effekten in Halbleitern, die der Wirkung mikroelektronischer Schaltkreise zugrunde liegen. Daß mit Hilfe elektronischer Schaltungen überhaupt Information verarbeitet werden kann, ist Gegenstand von Kapitel 4. In Kapitel 5 geht es um die Herstellungstechnologie der Mikroelektronik. Das Sortiment an mikroelektronischen Schaltkreisen, wie es die Industrie heute zur Verfügung stellt, wird in Kapitel 6 zunächst allgemein charakterisiert und danach an Hand ausgewählter Schaltkreise detailliert im Hinblick auf Funktion und Anwendung erläutert. Die künftige Entwicklung der Mikroelektronik, die wahrscheinlich weit über das heute Erreichte hinausführen wird, ist Gegenstand des letzten Kapitels. Das Buch überstreicht insgesamt also ein recht breites Spektrum von Fragen. Auf knappem Raum lassen sich diese Fragen nur dann behandeln, wenn ausgewählt und weggelassen wird. Das Prinzip, das wir dabei zugrunde legen, besteht darin, die für das Verständnis der Mikroelektronik entscheidenden Punkte ausführlich, verständlich, klar und möglichst vollständig darzustellen und andere Sachverhalte oft nur kurz oder überhaupt nicht abzuhandeln. Unser Anliegen ist es, die Mikroelektronik *begreiflich* zu machen;

wir streben nicht danach, über alles, was man zur Mikroelektronik wissen kann, möglichst umfassend zu informieren. Bei einem Fachlehrbuch sind mathematische Gleichungen und Ausdrücke ein wichtiges Hilfsmittel der Darstellung. Sie erleichtern die Verständigung zwischen Autor und Leser, sofern der Leser die mathematische Formelsprache auf dem erforderlichen Niveau beherrscht. Das konnte im vorliegenden Buch nicht vorausgesetzt werden. Auf die mathematische Formulierung der beschriebenen Sachverhalte haben wir daher weitgehend verzichtet. Wir bedienen uns der verbalen Erläuterung, die, wo immer das sinnvoll erschien, durch Abbildungen ergänzt wird. Das Buch ist dennoch keine leichte Lektüre. Trotz unserer Bemühungen um gute Verständlichkeit wird dem Leser manches recht kompliziert erscheinen. Innerhalb der gesteckten Grenzen wird er voll mit den Schwierigkeiten konfrontiert, die die Mikroelektronik dem Verständnis bietet. Diese Schwierigkeiten liegen in der Natur der Sache. Wer nicht an der Oberfläche bleiben will, muß sich ihnen unterziehen. Das Buch will stellenweise nicht gelesen, sondern studiert werden. Den Anstoß für dieses Buch haben öffentliche Vorlesungen gegeben, die der Autor in den zurückliegenden Jahren zu den physikalischen Grundlagen der Mikro- und Optoelektronik an der Humboldt-Universität zu Berlin gehalten hat. Beim Schreiben dieses Buches haben mich viele Fachkollegen unterstützt. Mein Dank gilt besonders Prof. H. Völz, Prof. J. Zaremba, Prof. E. Köhler, Dr. R. Strub, Dr. J. Dziesiaty und

Prof. H. Klose, die das Manuskript ganz oder teilweise gelesen und mich mit vielen wertvollen Ratschlägen unterstützt haben. Für nützliche Hinweise in Einzelfragen bedanke ich mich besonders bei Prof. J. Auth, Prof. H. Reimer, Dr. A. Rudolph, Dr. R. Behrendt, Prof. K.-H. Herrmann und Dr. H.-G. Schneider.

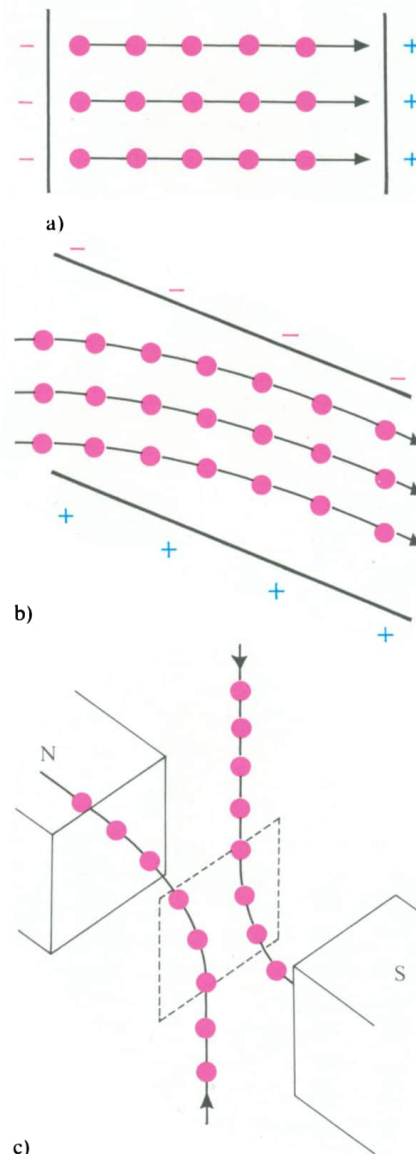
Zu danken habe ich auch dem Deutschen Verlag der Wissenschaften, dessen Leitung und Mitarbeiter mir bei der Herausgabe des Buches viel Verständnis entgegengebracht haben. Besonderen Dank schulde ich der Lektorin Frau Gisela Zielas und der Gestalterin Frau Gisela Deutsch für die sehr konstruktive Zusammenarbeit sowie Herrn Wolfgang Schwarzkopf und Herrn Hans Wunderlich für die sorgfältige Anfertigung der Zeichnungen.

Rolf Enderlein

1. Elektronik

Wir beginnen dieses Buch über Mikroelektronik mit einigen Erläuterungen zur *Elektronik* — ohne das Beiwort «Mikro». Worum geht es in der Elektronik? Elektronik hat etwas **Elektronen**

mit Elektronen zu tun. Elektronen sind sehr kleine, sehr leichte elektrisch geladene Elementarteilchen. Ihr Durchmesser beträgt $5 \cdot 10^{-13}$ cm, ihre Masse $9 \cdot 10^{-28}$ g und ihre Ladung $1,6 \cdot 10^{-19}$ Amperesekunden. Die Elektronen besitzen überhaupt die kleinste elektrische Ladung, die in der Natur anzutreffen ist, das Elementarquantum der elektrischen Ladung. Elektronen kommen in Atomen vor. Sie gehören neben den Protonen und Neutronen zu deren wesentlichsten Bestandteilen. Demzufolge sind Elektronen auch in allen größeren aus Atomen zusammengesetzten Systemen, wie festen Körpern, Flüssigkeiten oder Gasen, enthalten. Elektronen können aber auch als freie Teilchen auftreten. So wird der Kosmos neben anderen Elementarteilchen auch von Elektronen durchflogen. Aus einem radioaktiven Atomkern kommen beim sogenannten β -Zerfall Elektronen heraus. Feste Körper können durch Zuführung von Energie — z. B. in Form von Wärme, Licht oder durch Beschuß mit Ionen — veranlaßt werden, freie Elektronen abzugeben. Dieser Effekt wird bei einer Elek-



tronenröhre ausgenutzt. Die Katode einer solchen Röhre ist heizbar und gibt Elektronen ab, wenn sie erhitzt wird. Im freien Zustand halten sich Elektronen nur im Vakuum, d. h., wenn keine anderen Teilchen in der Nähe sind, an die sie sich anlagern können. Im Vakuum kann der Physiker mit den Elektronen verschiedene Experimente anstellen. Er kann sie mit Hilfe eines elektrischen Feldes auf eine hohe Geschwindigkeit bringen (Abb. 1 a). Er kann einen Strahl schon beschleunigter Elektronen durch Anlegen weiterer elektrischer Felder aus seiner ursprünglichen Richtung ablenken (Abb. 1 b). Indem er die Elektronen zusätzlich einem Magnetfeld aussetzt, kann er sie zwingen, sich auf Kreisbahnen zu bewegen (Abb. 1 c) usw. All das ist aber nicht nur mit den freien Elektronen im Vakuum möglich. Auch bestimmte Festkörper enthalten neben Elektronen, die stark an die Atome gebunden sind, relativ frei bewegliche Elektronen. Das gleiche trifft auf ionisierte Gase, sogenannte Plasmen, zu. Diese in Festkörpern oder Gasplasmen anzutreffenden beweglichen Elektronen lassen sich durch elektrische und magnetische

Abb. 1 Durch elektrische Felder werden Elektronen in Bewegung gesetzt (a) und abgelenkt (b). Magnetische Felder ändern die Bewegungsrichtung der Elektronen ebenfalls (c).

Felder ebenso leicht beeinflussen wie die Elektronen im Vakuum. Bei den Festkörperelektronen sind die Möglichkeiten der Beeinflussung sogar noch reichhaltiger. Außer auf elektrische und magnetische Felder reagieren diese Elektronen auch auf eine Bestrahlung des Festkörpers mit Licht, auf die Ausübung eines Druckes oder auf die Zuführung von Wärme, d. h. auf optische, mechanische oder thermische Einwirkungen. Es können sogar zusätzliche bewegliche Elektronen erzeugt werden, indem der Festkörper einer optischen oder thermischen Beeinflussung ausgesetzt wird. Wir haben damit bereits einen ersten allgemeinen Begriff davon erhalten, worum es in der Elektronik geht: Es geht um die Bewegung von Elektronen im Vakuum, im Festkörper, im Gasplasma und um die Beeinflussung dieser Bewegung bzw. der Zahl beweglicher Elektronen durch äußere Einwirkungen.

Die Beeinflussung geschieht dabei natürlich immer mit dem Ziel, bestimmte nützliche Zwecke zu erreichen. Auf welche Weise können Elektronen Nützliches leisten? Da ist zunächst einmal die Tatsache zu nennen, daß Elektronen überall dort im Spiele sind, wo Elektrizität praktisch verwendet wird. Die Elektronen sind die wichtigsten Träger der Elektrizität, wie sie im täglichen Leben vorkommt. Neben den Elektronen gibt es zwar auch noch elektrisch positiv geladene stabile Elementarteilchen, darunter insbesondere die *Protonen* und die *Positronen*. Aber die Positronen sind außerordentlich selten. Die Protonen sind zwar ebenso häufig wie die Elektronen und daher für die Elektrotechnik nicht ohne Be-

deutung, z. B. als Bestandteile von Ionen, wie sie in einem Blei-Akkumulator vorkommen. Ihr Nachteil ist aber, daß sie eine mindestens rund 2000 mal so große Masse wie die Elektronen haben und damit auch mindestens rund 2000 mal so schwerfällig sind wie diese. Für die meisten Anwendungen der Elektrizität sind deshalb die Elektronen besser geeignet als die Protonen. Wenn wir eine Glühlampe einschalten, bewegen sich Elektronen durch die Glühwendel, stoßen mit den Atomen zusammen und veranlassen diese, Licht auszusenden. Bei einem Elektromotor sind es letztlich die Kräfte auf die bewegten Elektronen, die den Anker in Rotation versetzen.

Elektrische

Energieumwandlung

Diese Leistungen der Elektronen lassen sich unter der Überschrift «Energieerzeugung» oder besser «*Energieumwandlung*» zusammenfassen. Elektronen können Energie tragen und umwandeln — elektrische Energie in Licht, Wärme, mechanische Energie oder chemische Energie. Sie sind die Hauptakteure, wenn elektrische Energie aus Wärme, chemischer Energie, Kernenergie, Sonnenenergie usw. erzeugt wird oder vom Ort ihrer Erzeugung zum Verbraucher transportiert wird. Wenn wir von Elektronik sprechen, haben wir gerade diese Fähigkeit der Elektronen zur Energieumwandlung nicht im Auge. Dabei werden gewissermaßen nur die allerelementarsten Eigenschaften der Elektronen ausgenutzt. Man bedient sich ihrer nur in einer relativ groben, undifferenzierten Art und Weise. Man nutzt lediglich, daß die Elektronen in Bewegung gesetzt werden können und,

einmal in Bewegung gesetzt, anderes zur Bewegung veranlassen können. Die Elektronik interessiert sich dagegen für feinere Effekte, nämlich primär dafür, wie eine bereits vorhandene Bewegung der Elektronen durch äußere Einwirkungen differenziert beeinflußt werden kann und welche unterschiedlichen Wirkungen auf diese Weise erreicht werden können. Der Gegenstand ist weniger die Bewegung der Elektronen selbst, als vielmehr deren gezielte Veränderung durch äußere Einflüsse. Daß dabei elektrische, magnetische, mechanische oder thermische Einflüsse in Frage kommen, haben wir bereits erwähnt. Wird der Strom in einem Stromkreis dadurch geändert, daß ein Widerstand erhitzt wird, so findet eine thermische Beeinflussung der Bewegung der Elektronen statt (Abb. 2a). Um eine elektrische Be-

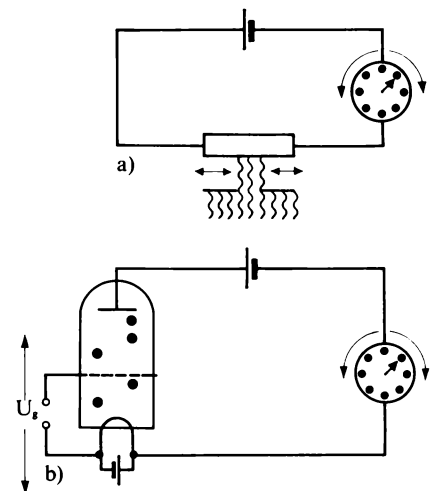


Abb. 2 Beeinflussung des Elektronenstroms
a) auf thermischem Wege durch unterschiedlich starke Beheizung eines Festkörperwiderstandes,
b) auf elektrischem Wege durch Veränderung der Gitterspannung U_g einer Elektronenröhre.

einflussung handelt es sich, wenn der Strom in einer Vakuum-Elektronenröhre durch Veränderung der Gitterspannung gesteuert wird (Abb. 2b). Obwohl die Elektronik auch andere Arten der Beeinflussung nutzt, ist die *elektrische* doch mit Abstand die wichtigste. Die Mikroelektronik, so wie wir sie in diesem Buch verstehen, verwendet ausschließlich die elektrische. In diesem Sinne können wir jetzt unsere Charakterisierung der Elektronik wie folgt präzisieren: In der Elektronik geht es darum, die Bewegung von Elektronen durch äußere *elektrische* Einwirkungen zu verändern und damit im Ergebnis bestimmte nützliche Zwecke zu erreichen.

Leistungselektronik und Informationselektronik

Wie wir bereits wissen, spielen Elektronen für die Energieumwandlung und -übertragung eine große Rolle. Werden Elektronen, die zum Zwecke der Umwandlung oder der Übertragung größerer Energiemengen in Bewegung gesetzt wurden, durch elektrische Einwirkungen gesteuert, so spricht man von *Leistungselektronik*. Anders ausgedrückt geht es in der Leistungselektronik um die elektrische Steuerung starker Ströme und großer Spannungen. Außer Energie können Elektronen aber noch etwas anderes tragen. Dieses andere ist *Information*. Ein Strom kann groß oder klein sein. «Der Strom ist groß» bedeutet eine Information. Auch die äußeren elektrischen Einwirkungen auf die Elektronen lassen sich in ihrer Stärke differenzieren. Zum Beispiel kann die Gitterspannung an der Elektronenröhre groß oder klein sein. Somit kann man also auch in die äußere

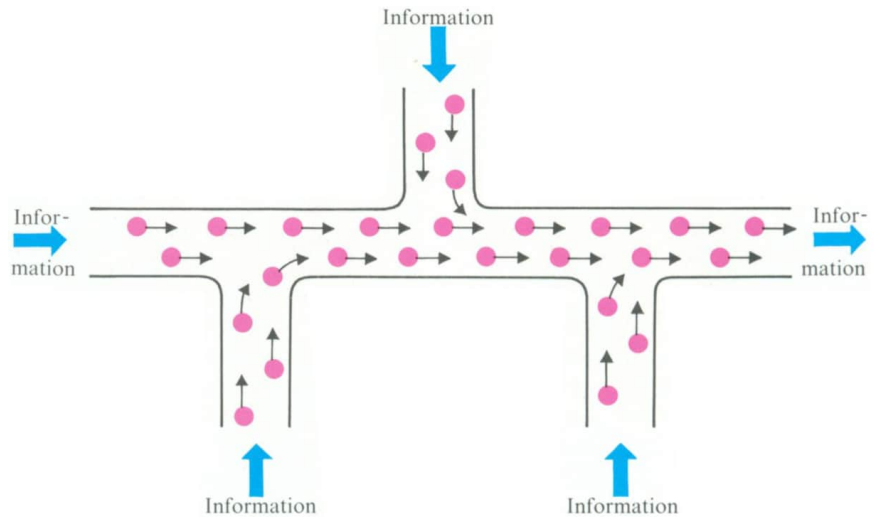


Abb. 3 Prinzip der Informationsverarbeitung mit Hilfe von Elektronen

Einwirkung eine Information hineinlegen. Bedenken wir noch, daß Elektronen selbst in der Lage sind, die Bewegung anderer Elektronen zu beeinflussen und damit also selbst die Rolle äußerer Einwirkungen übernehmen können, so gelangen wir zu folgender Feststellung: Elektronik ermöglicht das Übertragen von Information auf Elektronen, die Weitergabe dieser Information an andere Elektronen, d. h. die Verkopplung von Information, am Ende schließlich die Auslösung bestimmter mechanischer, optischer oder anderer dem Menschen direkt zugänglicher Wirkungen — mit einem Wort, *Verarbeitung von Information* (Abb. 3). Elektronik kann für die Verarbeitung von Informationen genutzt werden. Man spricht dann von *Informationselektronik*.

Wie die Verarbeitung von Information mit Hilfe der Elektronik im einzelnen geschieht, werden wir in den späteren Kapiteln noch ausführ-

lich behandeln. Im nächsten Kapitel wollen wir, ohne dabei schon auf technische Einzelheiten einzugehen, dem Leser zunächst einmal vor Augen führen, welche Rolle die Information und ihre Verarbeitung in der gesellschaftlichen Praxis spielen und welche Bedeutung dabei der Mikroelektronik zukommt.

2. Mikroelektronik, Information und Gesellschaft

Mikroelektronik ist Informations-elektronik. Mikroelektronik finden wir dort, wo etwas mit Information geschieht. Der Begriff «Information» ist dabei in einem sehr umfassenden Sinne gemeint. Wir verstehen darunter alle irgendwie gearteten Angaben oder Mitteilungen über Dinge und Erscheinungen. Information wird durch Rundfunk und Fernsehen verbreitet. Man kann Information aus der Zeitung entnehmen oder über das Telefon austauschen. Die Post ist eine Einrichtung, die es vorwiegend mit Information zu tun hat. Ein Taschenrechner operiert mit Information, und eine Uhr erzeugt sie laufend. Messen bedeutet Gewinnung von Information. Man benötigt Information, wenn Maschinen automatisch gesteuert werden sollen. Informationen sind nicht denkbar ohne einen Träger, d. h. ohne ein materielles Etwas, auf dem sie dargestellt werden können. Eine Nachricht existiert nicht an sich, sondern muß auf Papier gedruckt sein, durch Schwingungen der Luft übertragen werden oder eine andere materielle Gestalt besitzen. Auf dem gleichen Informationsträger kann Information in unterschiedlicher Weise dargestellt werden. Der Text auf dem Papier kann deutsch, russisch oder englisch oder überhaupt nicht mit Buchstaben, sondern in einem geheimen Zahlencode ge-

schrieben sein. Trotz unterschiedlicher Darstellung ist es immer die gleiche Information. Diese wenigen Bemerkungen zur Charakterisierung des Begriffes Information mögen zunächst genügen. Später werden wir noch einmal auf diesen Punkt zurückkommen und eine etwas genauere Definition geben.

Informationsprozesse in der Gesellschaft

Information ist nichts Statisches. Mit ihr geschieht in der Regel etwas, es finden *Informationsprozesse* statt. Informationsprozesse spielen im Leben der Gesellschaft und des Einzelnen eine ebenso große Rolle wie energetische und stoffliche Prozesse. Auch für das Leben im *biologischen* Sinne sind sie entscheidend. Im folgenden werden wir uns aber nur mit Informationsprozessen beschäftigen, die mit der Gesellschaft und der gesellschaftlichen Natur des Menschen verbunden sind, weil nur *sie* etwas mit Mikroelektronik zu tun haben. Man kann diese Prozesse in vier größere Gruppen einteilen: Erstens die Gewinnung, zweitens die Speicherung, drittens die Verarbeitung und viertens die Übertragung von Information.

Informationsgewinnung

Informationsgewinnung betreibt der Mensch zunächst einmal unmittelbar durch den Gebrauch seiner Sinnesorgane, vor allem des Auges, über

das der größte Teil der Information zu ihm gelangt. Aber die Leistungsfähigkeit der Sinnesorgane ist begrenzt, und viele Dinge und Erscheinungen sind ihnen überhaupt nicht direkt zugänglich. Deshalb spielen bei der Gewinnung von Information für den Menschen auch technische Hilfsmittel, vor allem Meßinstrumente, eine wichtige Rolle. Mit Hilfe von Meßinstrumenten vollzieht der Mensch eine aktive Informationsgewinnung über Objekte und Vorgänge in seiner Umwelt. Bei der medizinischen Diagnose ist der Mensch selbst das Meßobjekt. Informationen, die durch Messen gewonnen werden, sind aber nicht nur für den denkenden Menschen, sondern auch für Maschinen wichtig. Die dafür erforderlichen Meßfühler bezeichnet man in diesem Falle meist als *Sensoren*. Jedes Temperaturregelungsgerät besitzt einen Temperatursensor, für Warnanlagen werden Berührungs- oder optische Sensoren benötigt. Anspruchsvollere Industrieroboter sind mit Sensoren ausgerüstet. Wenn Informationen Angaben oder Mitteilungen über Dinge und Erscheinungen sind, so bedeutet Informationsgewinnung aber weit mehr als Messen. Jedes Erfassen oder Aufklären von Tatbeständen, jede Meinungsumfrage, jedes Beobachten und jede Erforschung der Vergangenheit ist in diesem Sinne Informa-

tionsgewinnung. Auch die Aufdeckung von Zusammenhängen in Natur und Gesellschaft oder die Schaffung von Kunstwerken muß man im weiteren Sinne dazu rechnen. Man erkennt daraus aber auch zugleich, daß die Grenzen zwischen der Informationsgewinnung und anderen Informationsprozessen, wie Speicherung oder Verarbeitung, fließend sind.

Informationsspeicherung

Wird die Information nicht nur für den Augenblick ihrer Gewinnung, sondern auch noch später benötigt, oder wird die Information überhaupt nur für eine spätere Verwendung produziert, so macht sich eine *Speicherung* erforderlich. Eine auf Magnetband aufgezeichnete Life-Sendung ist eine Information von der ersten Art, der Inhalt eines Buches eine von der zweiten Art. Die Notwendigkeit zur Speicherung ergibt sich für Informationen aus allen Bereichen des Lebens. Es kann sich dabei um wissenschaftliche Daten handeln, um Konstruktionsunterlagen, um Angaben über aktuelle Ereignisse oder um Mitteilungen über vergangene Zeiten. Medizinische Diagnoseergebnisse sind ebenso zu speichern wie Informationen über den Zustand nationaler Ökonomien oder privater Kontostände. Der Zweck der Informationsspeicherung ist es, vorhandene Informationen über einen längeren oder kürzeren Zeitraum sicher und gut zugänglich aufzubewahren. Als Speicher werden heute Bücher, Zeitschriften, Zeitungen, Akten, Fotografien, Mikrofilme, Lochkarten, Schallplatten, Magnetbänder, Disketten, elektronische Schaltkreise und vieles andere mehr verwendet. In Bibliotheken, Bü-

ros, Rechenzentren oder Datenbanken werden die gespeicherten Informationen gesammelt. Die Anforderungen an die Informationsspeicherung sind mit fortschreitender Entwicklung der Gesellschaft ständig gewachsen. Es wurde und wird viel mehr neue speichernswerte Information gewonnen, als bereits gespeicherte dem Vergessen überantwortet werden kann. Mit der ständig wachsenden Informationsflut fertigzuwerden ist heute zu einem großen Problem geworden. Es besteht nicht nur darin, auf begrenztem Raum immer mehr Information zu speichern, sondern vor allem auch darin, die gespeicherte Information besser handhabbar zu machen. Das bedeutet unter anderem, den schnellen Zugriff zur gewünschten Information auch dann zu gewährleisten, wenn man nicht weiß, an welcher Stelle im Speicher sie untergebracht ist, sondern nur allgemeine Angaben über sie machen kann. In der Arbeit des Wissenschaftlers tritt dieser Fall sehr häufig auf. Man möchte wissen, welche Ergebnisse auf einem Gebiet, auf dem man selbst tätig ist, bereits publiziert worden sind. Daß dieses Problem heute nur unvollkommen gelöst ist, beweisen u. a. Doppelarbeiten in der Wissenschaft.

Informationsverarbeitung

Die Speicherung ist ein Informationsprozeß, der vorhandene Information nicht verändert. Es gibt andere Prozesse, die vorhandene Information in neue, vorher nicht vorhandene Information umwandeln. Man faßt diese Prozesse unter der Bezeichnung *Informationsverarbeitung* zusammen. Die Ausführung einer Rechenaufgabe mit vorgegebenen Zahlen und Rechen-

schriften ist ein typisches Beispiel für die Informationsverarbeitung. Ist die Rechnung komplizierter, so liegt ein Problem für die *Rechentechnik* vor. Handelt es sich um eine einfache Rechnung, die aber mit einem umfangreichen Zahlenmaterial durchzuführen ist, spricht man von *Datenverarbeitung*. Allgemein werden bei der Datenverarbeitung nicht nur Zahlen verwendet, sondern auch Buchstaben und Zeichen und außer dem Rechnen auch noch andere Verarbeitungsschritte. Die Datenverarbeitung besitzt deshalb so große Bedeutung, weil sich viele Informationen relativ einfach mit Hilfe von Zahlen, Buchstaben und Zeichen darstellen lassen. Zum Beispiel kann man Bücher und Ausleiher oder Fluglinien, Flugzeuge und Passagiere durch Nummern und Buchstaben kennzeichnen. Auf diese Weise wird die Arbeit mit der Kartei einer Bibliothek oder mit den Buchungslisten im internationalen Flugverkehr zu einer Aufgabe der Datenverarbeitung. Liegt die zu verarbeitende Information in Form eines Textes vor, spricht man von *Textverarbeitung*. Sie kann darin bestehen, daß ein Text korrigiert oder überarbeitet werden muß, daß eine bestimmte Textstelle aufzufinden oder eine Auswertung unter einem gegebenen Gesichtspunkt vorzunehmen ist. Neben Daten und Texten müssen in der Praxis auch Bilder verarbeitet werden. Man spricht dann von *Bildverarbeitung*. Ein Beispiel dafür ist die Erkennung von Zeichen oder Bildern. Dazu muß überprüft werden, mit welchem Exemplar aus einer Auswahl verschiedener Vorlagen sie übereinstimmen. Dieses Problem tritt zum Beispiel bei Industrierobotern auf, die

erkennen sollen, ob ein zu bearbeitendes Werkstück die richtige Form und Lage hat. Oft hat Information auch die Form *gesprochener Wörter*. Für die Informationsabgabe durch den Menschen hat das Sprechen sogar eine ganz besondere Bedeutung. Durch keine andere physische Aktivität kann der Mensch so viel Information abgeben wie durch das Sprechen, vorausgesetzt, er verfügt nicht über spezifische Hilfsmittel wie etwa eine schnelle Schreibmaschine und dazu über ein spezielles Training im Maschine-Schnellschreiben. Mit dieser Einschränkung ist das Sprechen derjenige Informationskanal, über den der Mensch pro Zeiteinheit die größte Informationsmenge nach außen übermitteln kann, weit mehr als etwa durch das Hervorbringen optisch sichtbarer Zeichen oder Bilder. Gerade dadurch hat sich die Sprache zum wichtigsten Mittel für den Informationsaustausch zwischen den Menschen entwickelt. Daraus ergeben sich auch Konsequenzen für die Kommunikation des Menschen mit Maschinen und technischen Anlagen. Es wäre günstig, wenn auch dabei der Informationsaustausch unmittelbar über gesprochene Wörter erfolgen könnte. Damit das möglich ist, müssen Maschinen einerseits Sprache verstehen und andererseits auch hervorbringen können. Spracherkennung und Sprachsynthese werden damit, ähnlich wie die Daten-, Text- und Bildverarbeitung, zu praktisch bedeutungsvollen Problemen der Informationsverarbeitung. Industrieroboter, die Bilder erfassen können, und Maschinen, die Sprache verstehen, sind Beispiele aus einem großen Bereich, in dem Information

eine wichtige Funktion hat — dem Bereich der materiellen Produktion. Die Information spielt hier die Rolle des unsichtbaren Lenkers, entweder für den Menschen und seine Einwirkungen auf die Maschinen oder auch direkt für die Maschinen, ohne daß der Mensch dazwischen tritt. Produktionsprozesse sind *auch* Informationsprozesse. Die Bearbeitung eines Werkstücks auf einer Werkzeugmaschine ist außer einem mechanischen Vorgang auch ein Prozeß, bei dem Information verarbeitet werden muß. Der mechanische Bearbeitungsvorgang muß so ablaufen, daß am Ende das geforderte Werkstück herauskommt. Dazu muß während eines bestimmten Bearbeitungsschrittes, etwa dem Drehen, gemessen und die Bearbeitung fortgesetzt werden, wenn die geforderten Maße noch nicht erreicht sind. Der folgende Bearbeitungsschritt, etwa das Bohren, muß eingeleitet werden, wenn sie erreicht sind usw. Es ist also notwendig, Informationen durch Messen zu gewinnen, diese zu verarbeiten und in Handlungen umzusetzen. In der chemischen Industrie werden Stoffe umgewandelt, z. B. Steinsalz, Ammoniak und Kohlendioxid in Soda für die Glas- und Waschmittelindustrie oder Azetylen und Salzsäure in Polyvinylchlorid (PVC) für die Herstellung von Plasterzeugnissen. Damit die chemischen Reaktionen mit einer hohen Ausbeute an qualitätsgerechten Endprodukten bei minimalem Energie- und Rohstoffeinsatz ablaufen können, sind in den Reaktoren die für den Prozeß optimalen Bedingungen einzustellen. Die Temperatur muß auf dem günstigsten Wert gehalten und bei Abweichungen von diesem muß

Wärme zu- oder abgeführt werden. Der Druck oder die Konzentration der beteiligten Stoffe müssen überwacht und nötigenfalls an die optimalen Werte herangeführt werden. All das ist nicht nur an *einer* Stelle, sondern an vielen Stellen der chemischen Anlage erforderlich. Das, was hier über den Betrieb einer chemischen Anlage und über die Bearbeitung eines Werkstücks gesagt wurde, gilt für jeden Arbeitsvorgang in der materiellen Produktion. Jeder Arbeitsvorgang ist neben der unmittelbaren physikalischen oder chemischen Einwirkung auf den Arbeitsgegenstand gleichzeitig auch ein Prozeß der Informationsverarbeitung. Diese Feststellung, die hier zunächst eingeschränkt auf die materielle Produktion getroffen wurde, behält ihre Gültigkeit auch dann, wenn wir den Begriff «Arbeit» in einem viel weiteren Sinne auffassen. Nicht nur die unmittelbaren Herstellungsschritte erfordern Arbeit, sondern auch alle vor- und nachgelagerten Prozesse, beginnend mit der Entwicklung des Produktes im Forschungslaboratorium, der Anfertigung der Konstruktionsunterlagen und der Materialbereitstellung über die Qualitätskontrolle am fertigen Erzeugnis, seine Verpackung und Lagerung bis hin zur Auslieferung. Es ist nicht übertrieben, wenn wir feststellen, daß bei vielen Produkten diese Vor- und Nacharbeiten wesentlich mehr Zeit in Anspruch nehmen als der eigentliche Herstellungsprozeß. Entsprechend umfangreich ist die Information, die dabei verarbeitet werden muß. Schließlich ist die materielle Produktion auch nicht ohne Verwaltung, Planung und Leitung denkbar — Prozesse, die außer

für die Produktion natürlich auch für alle anderen gesellschaftlichen Bereiche eine große Rolle spielen und die in vielfältiger Weise mit Informationsprozessen verbunden sind, ja, die nicht selten *selbst* Informationsprozesse sind. Ein anderer Bereich menschlicher Tätigkeit, in dem Voraussetzungen für die materielle Produktion geschaffen werden, ist die Wissenschaft. Daß die Gewinnung und Verarbeitung von Information gerade in der Wissenschaft eine große Rolle spielen, braucht nicht weiter erläutert zu werden, genauso wenig wie betont werden muß, daß es an Schulen und Universitäten oder ganz allgemein im Bildungswesen neben anderem um die Vermittlung von Information geht.

Zu den der Produktion nachgelagerten Prozessen gehören im weitesten Sinne auch der Transport der Waren und der Verkauf durch den Handel. Auch dabei sind überall Informationsprozesse im Spiel. Der Lieferwagenfahrer, der sein Auto durch den Verkehr steuert, hat ständig Informationen über die Verkehrslage, über den momentanen Zustand seines Fahrzeuges und über die Beschaffenheit der Fahrbahn in sich aufzunehmen, zu verarbeiten und, geleitet von den Verkehrsregeln, Bedienungsoperationen auszuführen. Auch mir selbst ergeht es nicht anders, wenn ich mit meinem Auto fahre. Ist das Ziel meiner Fahrt ein Kaufhaus, und ist es meine Absicht, etwas einzukaufen, so löse ich eine ganze Flut weiterer Informationsverarbeitungsprozesse aus. Möglicherweise wird bereits mein Eintreten in das Kaufhaus von einer Lichtschranke registriert, damit die Di-

rektion ständig über die Zahl der Kunden im Hause informiert ist. Dadurch, daß ich die gewünschten Gegenstände kaufe, verändere ich die Bestände des Kaufhauses an diesen Waren. Der aktuelle Warenbestand ist eine wichtige Information, die ebenfalls durch das Kaufhaus erfaßt werden muß. Die Registrierkasse bei der Bezahlung der zu kaufenden Gegenstände gibt mir die Information, wieviel ich insgesamt zu bezahlen habe, und verfolgt gleichzeitig für das Kaufhaus die Entwicklung der Einnahmen. Zahle ich mit einem Scheck oder einer Kreditkarte, so informiere ich damit die *Bank* über den Auftrag, von meinem Konto die bezeichnete Summe abzubuchen und dem Konto des Kaufhauses gutzuschreiben. Die Höhe meines Kontos auf der Bank ist im Prinzip auch nichts weiter als eine in «so und so viel Mark» ausgedrückte Information. Diese «so und so viel Mark» sind für mich nicht in Papier oder Münzen, geschweige denn in einem Wertäquivalent wie etwa dem Gold, in einem Kästchen mit meiner Kontonummer in der Bank deponiert, sondern die entsprechende Zahl ist unter meiner Kontonummer auf Papier aufgeschrieben oder auf einem Magnetband gespeichert. Monatlich wird die Bank durch meine Arbeitsstelle darüber informiert, daß zu dieser Summe eine weitere Summe, mein Gehalt, zu addieren ist. Die Höhe meines Gehaltes ist letztlich nichts weiter als die Information darüber, in welchem Umfang ich aufgrund meiner Arbeitsleistung berechtigt bin, Leistungen anderer für mich in Anspruch zu nehmen. Im *Zahlungsverkehr* werden Informationen dar-

über ausgetauscht, wie sich die Ansprüche auf Leistungen anderer zwischen den beteiligten Partnern verändern. Also ist auch der Zahlungsverkehr ein Informationsprozeß, und man braucht sich nicht darüber zu wundern, wenn dafür heute bereits die Mittel elektronischer Informationstechnik eingesetzt werden. Dabei handelt es sich allerdings noch um Experimente, wie etwa dem in der französischen Provinzstadt Biarritz. Es muß sich erst noch zeigen, wie weit das elektronische Zahlungsverfahren dem herkömmlichen tatsächlich überlegen ist. Daß im Zahlungsverkehr letztlich Information bewegt wird, folgt jedoch schon aus der Existenz solcher Experimente.

Informationsübertragung

Für die Rolle, die die *Informationsübertragung* in der Gesellschaft spielt, ist der Zahlungsverkehr nur ein kleines Beispiel am Rande. Der Austausch von Informationen, die *Kommunikation*, ist eine der entscheidenden Entwicklungs- und Existenzbedingungen der menschlichen Gesellschaft überhaupt. Die Fähigkeit zur Kommunikation mit seinesgleichen besitzt der Mensch schon von Natur aus. Aber sie reicht nur so weit, wie das Auge sehen und das Ohr hören kann. Für den Informationsaustausch über größere Entfernungen, die *Telekommunikation*, ist der Mensch von Natur aus nicht eingerichtet. Dafür benötigt er entweder lebende Boten oder technische Hilfsmittel. Bis in das vergangene Jahrhundert hinein beschränkten sich die Hilfsmittel auf Trommeln, Glocken, Rauchzeichen, starke Lichtquellen u. ä. Die Wende kam mit der Entdeckung der Elektrizität.

tät und der elektromagnetischen Wellen. Damit standen technische Hilfsmittel zur Verfügung, die sich für die Telekommunikation als außerordentlich leistungsfähig erwiesen, so leistungsfähig, daß ihre Potenzen bis heute bei weitem noch nicht ausgeschöpft sind. Das ist auch gut so, denn die Entwicklung der Gesellschaft erfordert, daß immer mehr Information über große Entfernungen übertragen wird. Technische Einrichtungen, die dafür heute zur Verfügung stehen, sind unter anderem Fernschreiber und Faksimilegeräte für die Übermittlung von Zeichen, Texten und Bildern oder das Telefon für die Übertragung von Sprache und in Zukunft in zunehmendem Maße wahrscheinlich ebenfalls von Bildern. Gerichtete Funkverbindungen ermöglichen die Kommunikation mit Raumschiffen und die Nachrichtenübertragung mit Hilfe von Satelliten. Rundfunk und Fernsehen dienen ebenfalls der Übertragung von Information, wozu in dem breiten Sinne, in dem wir Information hier verstehen, auch künstlerische Darbietungen und Unterhaltung gerechnet werden müssen. Neben den elektronischen Massenmedien Rundfunk und Fernsehen wird aber nach wie vor auch durch nichtelektronische Medien wie Zeitungen, Bücher oder Kinofilme Information in großem Umfang verbreitet.

Informationsbegriff

Nachdem wir an einer Reihe typischer Beispiele kennengelernt haben, was mit Information, ihrer Gewinnung, Speicherung, Verarbeitung und Übertragung gemeint ist, wollen wir uns noch einmal dem *Begriff der Information* allgemein zuwenden

und eine etwas genauere Definition versuchen, als sie bisher erfolgt ist. Weiter oben hatten wir davon gesprochen, daß wir unter Information alle irgendwie gearteten Angaben oder Mitteilungen über Dinge und Erscheinungen verstehen wollen. Diese Definition besagt jedoch nicht sehr viel, weil wir im Grunde nur das Wort «Information» durch die Wörter «Angaben» und «Mitteilungen» ersetzt haben. Wir kommen dem Wesen der Sache näher, wenn wir nicht fragen, was Information ist, sondern was sie *bewirkt*. Information bewirkt die Beseitigung von Ungewißheit. Information setzt uns oder eine technische Anlage in den Stand, zwischen einer Anzahl mehr oder weniger gleichberechtigter Möglichkeiten auszuwählen. Information kommt also dann ins Spiel, wenn erstens ein bestimmter Vorrat an Möglichkeiten besteht und wenn zweitens eine bestimmte Bewertung dieser Möglichkeiten vorgenommen werden muß. Information verleiht uns oder der technischen Anlage die Fähigkeit, die geforderte Bewertung vorzunehmen. Dafür, wie das Wetter heute ist, gibt es eine ganze Palette von Möglichkeiten. Es kann regnen oder windig sein, es kann die Sonne scheinen, der Himmel kann bedeckt sein ohne daß es regnet, die Temperatur kann 0 °C, 10 °C oder 20 °C betragen. Die Angabe, daß das Thermometer heute 10 °C zeigt, bewertet Varianten für das Wetter mit einer davon abweichenden Temperatur als nicht zutreffend, alle anderen Varianten als möglich. Ein Wanderer, der sich auf dem Weg zu einem bestimmten Ziel befindet, wird, wenn er an eine Weggabelung gerät, plötzlich mit zwei Möglichkeiten

weiterzugehen konfrontiert. Er hat sich für *den* Weg zu entscheiden, der ihn zum Ziel führt. Wenn er den Weg bereits kennt, benötigt er dafür keine Information. Er ist von vornherein in der Lage, die ihm abverlangte Bewertung der beiden Möglichkeiten vorzunehmen. Ist die Situation für ihn aber neu und hat er vorher keinerlei Anhaltspunkte zur Bewertung der beiden Wege, so braucht er eine Information. Man ist übereingekommen, die Menge an Information, die man bei völliger vorheriger Unkenntnis zur Entscheidung zwischen zwei gleichberechtigten, gänzlich voneinander unabhängigen Möglichkeiten benötigt, als *Maßeinheit der Information* anzusehen und als *bit* zu bezeichnen. Um unter vier verschiedenen Wegen den richtigen auswählen zu können, benötigt man 2 bit Information. Das kann man sich folgendermaßen klar machen. Man ordnet die vier Wege zunächst in zwei Gruppen zu je zwei Wegen. Die Entscheidung für die Gruppe mit dem richtigen Weg erfordert, wie jede Auswahl zwischen zwei gleichberechtigten Möglichkeiten, 1 bit Information. Nachdem die Gruppe bekannt ist, in der sich der richtige Weg befindet, muß unter den beiden Wegen dieser Gruppe der richtige ausgewählt werden. Dafür wird noch einmal 1 bit Information benötigt. Insgesamt sind es also 2 bit. Hätten wir acht Möglichkeiten zur Auswahl gestellt, wären es 3 bit gewesen, bei 16 Möglichkeiten 4 bit usw. Man erkennt daraus bereits die allgemeine Gesetzmäßigkeit. Mit n bit Information kann man unter 2^n Möglichkeiten die gesuchte finden. Umgekehrt bedeutet die Angabe der einen richtigen Möglichkeit

unter 2^n gleichermaßen in Frage kommenden eine Information von n bit. Die Information ist also der Logarithmus zur Basis 2 ($\log_2$) aus der Anzahl der Möglichkeiten, denn nach der Definition des Logarithmus zur Basis 2 ist $n = \log_2 2^n$. In diesem Sinne können wir folgende Gleichung schreiben:

Information = $\log_2$ aus der Anzahl der Möglichkeiten.

Damit ist eine exakte Vorschrift zur Bestimmung des Informationsgehaltes von Angaben und Mitteilungen gegeben. Sie geht auf den amerikanischen Mathematiker Claude Shannon zurück, der sie im Zusammenhang mit seinen Untersuchungen zur Nachrichtenübertragung 1947 aufgestellt und verwendet hat. Der Shannonsche Begriff der Information ist eng mit dem thermodynamischen Begriff der Entropie verwandt, der von einem deutschen und einem österreichischen Physiker, Rudolf Clausius und Ludwig Boltzmann, bereits in der zweiten Hälfte des 19. Jahrhunderts begründet wurde. Die Shannonsche Gleichung kann auch dann angewandt werden, wenn die Anzahl der Möglichkeiten keine Potenz von zwei ist, wie das bisher der Fall war. Um etwa unter drei Wegen den richtigen auswählen zu können, ist eine Information von $\log_2 3 = 1,58$ bit erforderlich. Jeder der insgesamt 26 Buchstaben des lateinischen Alphabets enthält $\log_2 26 = 4,70$ bit Information. Beim Schreiben eines Textes benötigt man außer den Buchstaben noch Satzzeichen und die Trennungslücke, zusammen mit den Buchstaben also etwa $2^5 = 32$ Zeichen. Jedes Zeichen des Textes hat auf diese Weise 5 bit Information.

Ein Buch mit 100 Seiten und 2000 Zeichen pro Seite enthält also 10^6 bit Information. Diese Aussage ist so zu verstehen, daß wir eine Informationsmenge von 10^6 bit aufnehmen, wenn wir uns beim Lesen überzeugen, daß die Verteilung der 32 Buchstaben und Zeichen im Buch nicht irgendwie, sondern gerade so erfolgt ist, wie wir sie vor Augen sehen. Beim Lesen des Buches erhalten wir die Information, daß es nicht irgendeines der insgesamt 10^6 möglichen verschiedenen Bücher mit 100 Seiten ist, sondern gerade das vorliegende. Diese Berechnung des Informationsgehaltes eines Buches mutet allerdings etwas merkwürdig an. Sie ist es in der Tat auch. Nach dieser Berechnung enthält auch ein völlig unverständlicher oder unsinniger Text, z. B. einer, der nur aus einzelnen «e» oder aus Kommas besteht, also ein Text ohne jeglichen Informationswert, 10^6 bit Information. Der Grund für dieses merkwürdige Ergebnis ist die Shannonsche Definition, die nicht nach dem Wert oder der Bedeutung der Information fragt. Sie charakterisiert Information nur als etwas, das die Ungewißheit angesichts eines Vorrats an gleichberechtigten Möglichkeiten einschränkt oder beseitigt. Unser Gefühl sagt uns aber, daß Information darüber hinaus noch etwas mehr ist. Wir bewerten Information auch danach, ob sie interessant, wichtig oder nützlich ist, also nach Merkmalen, die in der Shannonschen Definition überhaupt nicht vorkommen. Durch diese Definition werden offensichtlich nicht alle Seiten der Information erfaßt, sondern es wird nur ihre mathematisch-statistische Seite berücksichtigt. Die

Frage, ob eine Verallgemeinerung des Informationsbegriffs möglich ist, die dem Wesen der Information noch besser entspricht, ist vollaufberechtigt und wird heute unter Fachleuten diskutiert. Allerdings fehlen bisher abschließende Antworten. Trotz seiner Unvollkommenheit hat sich der Shannonsche Informationsbegriff als wirkungsvolles Instrument erwiesen. Er wird heute überall in Wissenschaft und Technik und darüber hinaus mit Erfolg verwendet.

Information und Biologie

Fast alles, was wir bisher über Information gesagt haben, bezog sich auf die menschliche Gesellschaft. Daneben gibt es einen zweiten großen Realitätsbereich, in dem Information eine wichtige Rolle spielt. Dieser Bereich ist die *Biologie*. Informationsprozesse sind entscheidende Bedingungen für die Existenz von Lebewesen und für deren Fortbestehen und Höherentwicklung in den Nachkommen. In den Sinnesorganen und dem Nervensystem verfügen alle entwickelteren Lebewesen über ein eigenständiges System zur Gewinnung, Verarbeitung und Speicherung von Information. Das Gedächtnis des Menschen besitzt eine Speicherkapazität, die wahrscheinlich oberhalb von 10^{10} bit liegt. Auch die Vererbung beruht ganz auf der Speicherung und Übertragung von Information. Die in den Molekülen der Desoxyribonukleinsäure (DNS) speicherbare Erbinformation beträgt beim Menschen grob gerechnet 10^{10} bit. Das ist so viel Information, wie in einigen 100 Büchern zu je 1000 Seiten untergebracht werden kann. In jeder einzelnen Zelle eines Lebewesens, mit

Ausnahme der Keimzellen, ist die Erbinformation wahrscheinlich vollständig oder fast vollständig enthalten. Darin kommt einerseits die außerordentlich kompakte, platzsparende Art und Weise zum Ausdruck, in der die Natur Information speichern kann. Sie ist der Mikroelektronik in dieser Beziehung haushoch überlegen. Andererseits ist der geradezu verschwenderische Umgang mit Information bemerkenswert — eine einzige Zelle mit der eingeschriebenen Information würde bereits ausreichen, alles darüber hinaus ist Redundanz. Aber diese Redundanz bedeutet Sicherheit, und zwar nicht 5 oder 10fache, wie wir sie aus der Technik kennen, sondern millionenfache. Durch die Erzeugung von Kopien der DNS-Moleküle wird die Erbinformation an die Nachkommen weitergegeben. Mitunter kommt es in den Originalen zu spontanen Änderungen, sog. Mutationen. Auch sie werden auf die Nachkommen übertragen. Durch die natürliche Auslese setzen sich die leistungsfähigeren Individuen gegenüber den weniger tüchtigen durch. Es findet das statt, was man als biologische Evolution bezeichnet. Auch dabei spielt die Information eine entscheidende Rolle. Informationsprozesse im biologischen Bereich haben wenigstens in zweierlei Hinsicht auch Beziehung zur Technik. Erstens kann die Technik von den Lösungen lernen, die die Natur für die Informationsprozesse in den Lebewesen gefunden hat. Zweitens kann sie versuchen, diese Prozesse in einer gewünschten Richtung zu beeinflussen, wie das in der Gentechnologie heute bereits geschieht.

Die Informationsprozesse, die Gegenstand der Mikroelektronik sind, haben alle mit der Gesellschaft und mit der gesellschaftlichen Natur des Menschen zu tun. Kehren wir also nach diesem Abstecher in die Biologie zu den Informationsprozessen in der Gesellschaft zurück. Es gibt sie in einer unüberschaubaren Fülle und Vielgestaltigkeit. Sie kommen in allen Bereichen der Gesellschaft vor — in der materiellen Produktion, im Transportwesen, in der Verwaltung, bei der Planung und Leitung gesellschaftlicher Prozesse, in der Kommunikationssphäre, in der Wissenschaft, der medizinischen Betreuung, im Haushalt und in der Freizeit der Menschen. Es gibt keinen praktisch bedeutungsvollen Vorgang im Leben der Gesellschaft oder des Einzelnen, der nicht zugleich auch ein Informationsprozeß wäre. Die Informationsverarbeitung spielt dabei insofern eine besondere Rolle, als einerseits alle übrigen Prozesse, also die Gewinnung, Speicherung und Übertragung von Information, meist ohne Verarbeitung nicht denkbar sind und weil andererseits die übrigen Informationsprozesse in vielen Fällen nur deshalb durchgeführt werden, weil sie für die Informationsverarbeitung als eigentlich interessierender Prozeß benötigt werden. In diesem Sinne ist die Informationsverarbeitung der universellste Informationsprozeß. Die Elektronik ist in der Lage, jedes **Informationsverarbeitung mit Hilfe der Mikroelektronik** Problem der Informationsverarbeitung zu lösen, vorausgesetzt, daß es in eine mathematische Form gebracht und ein Schema zu seiner Lösung, ein sogenannter Algorithmus,

angegeben werden kann. Viele der in der Praxis anfallenden Informationsprobleme erfüllen diese Forderung von vornherein. Es gibt aber auch andere, die als Ganzes einer mathematischen Lösung entweder prinzipiell nicht oder heute noch nicht zugänglich sind. In vielen Fällen lassen sich aber auch aus solchen Problemen algorithmisierbare Teilaufgaben ableiten. Man denke etwa an die Aufstellung des Finanzplanes für einen privaten Haushalt oder an die Formulierung des Forschungsprogramms für ein Institut. Für die Entscheidungen, die in beiden Fällen getroffen werden müssen, spielen Wertvorstellungen, Situationsbeurteilungen und Zukunftserwartungen, d. h. Kategorien eine Rolle, die sich nicht in mathematischer Form ausdrücken lassen. Es gibt deshalb für keines der beiden Probleme einen vollständigen Lösungsalgorithmus. Dennoch kommen in beiden Fällen auch algorithmisierbare Teilschritte vor. Im folgenden beziehen wir uns stets auf Informationsverarbeitungsprobleme, für die es einen mathematischen Lösungsalgorithmus gibt. Die Menge dieser Probleme umfaßt zwar nicht alles, was man im weitesten Sinne unter Informationsverarbeitung verstehen kann, aber sie ist immer noch unvorstellbar groß. Mit Hilfe der Elektronik kann jedes beliebige dieser Probleme gelöst werden. Das ist letztlich der Ursprung ihrer großen Bedeutung. Die Elektronik bietet für die in unüberschaubarer Fülle und Vielgestaltigkeit auftretenden algorithmisierbaren Probleme der Informationsverarbeitung *prinzipiell* eine Lösung. Die Betonung liegt hier auf *prinzipiell*. Die Mikroelektronik macht die prinzi-

Tab. 1 Leistungsbestimmende Kenngrößen der Einzel-Bauelemente und der zentralen Verarbeitungseinheit eines Rechners mit 10^4 Bauelementen auf den verschiedenen Stufen der Elektronik. Die Zahlenwerte stellen nur eine grobe Orientierung dar

		Volumen cm ³	Leistungsverbrauch Watt	Arbeitsgeschwindigkeit Schaltakte bzw. Rechenschritte pro Sekunde	Ausfallrate pro Stunde	Preis M
1 Bauelement	Röhre	10 ¹	10 ⁰	10 ⁸	10 ⁻⁵	10 ¹
	Einzeltransistor	10 ⁻²	10 ⁻¹	10 ⁸	10 ⁻⁸	10 ⁰
	Integrierter Transistor (Mikroelektronik)	10 ⁻⁸	10 ⁻⁵	10 ⁸	10 ⁻¹¹	10 ⁻⁴
Rechner mit 10 ⁴ Bauelementen	Röhren	10 ⁷	10 ⁴	10 ⁶	10 ⁻¹	10 ⁶
	Einzeltransistor	10 ⁴	10 ¹	10 ⁶	10 ⁻⁴	10 ⁵
	Schaltkreise (Mikroelektronik)	10 ⁻²	10 ⁻¹	10 ⁶	10 ⁻⁷	10 ¹

piell mögliche Lösung in vielen Fällen auch *praktisch realisierbar*. Praktisch realisierbar heißt: mit vertretbaren Kosten, mit geringem Aufwand an Energie und Material, mit der erforderlichen Zuverlässigkeit und Schnelligkeit. Der Unterschied zwischen «prinzipiell möglich» und «praktisch realisierbar» muß noch etwas genauer beleuchtet werden, weil er den entscheidenden Ansatzpunkt für die Entwicklung der Mikroelektronik darstellt. 1945 wurde eine der ersten elektronischen Rechenmaschinen, die ENIAC, in Betrieb genommen. Sie enthielt ca. 18000 Elektronenröhren. Das war bis dahin mit Abstand die größte Zahl aktiver Bauelemente, die jemals in einem elektronischen Gerät verbaut worden war. Dieser Rechner leistete aber trotzdem nur soviel wie heute ein Taschenrechner. Vergewegenwärtigen wir uns einmal (siehe Tabelle 1), welche Größe, welcher Energieverbrauch, welche Schnelligkeit, welche Zuverlässigkeit und welcher Preis sich für die

zentrale Verarbeitungseinheit eines solchen Rechners ergeben, wenn die verschiedenen Realisierungsmöglichkeiten der Elektronik angewandt werden, nämlich erstens Elektronenröhren, zweitens die diskreten Halbleiterbauelemente und drittens die integrierten Schaltkreise der Mikroelektronik. Die Tabelle beweist: für den massenhaften Einsatz realisierbar ist ein solcher Rechner erst mit Hilfe der Mikroelektronik. Dabei ist der extrem geringe Platzbedarf der zentralen Verarbeitungseinheit des Rechners bei der mikroelektronischen Lösung mit $10^{-2} \text{ cm}^3 \approx 4 \text{ mm} \times 4 \text{ mm} \times 0,5 \text{ mm}$ zwar von Vorteil, aber auch die $10^4 \text{ cm}^3 \approx 40 \text{ cm} \times 40 \text{ cm} \times 5 \text{ cm}$ bei Verwendung von diskreten Halbleiterbauelementen wären noch einigermaßen handlich. Die entscheidenden Vorteile der mikroelektronischen Lösung sind die geringen Kosten, die hohe Zuverlässigkeit und der geringe Energieverbrauch bei gleichzeitig hoher Arbeitsgeschwindigkeit. Notwendige Voraussetzung für diese Vorteile ist

die extreme Kleinheit der integrierten Bauelemente. Die Miniaturisierung der Bauelemente in der Mikroelektronik ist demnach weniger Selbstzweck als Mittel zum Zweck. Der Zweck sind letztlich elektronische Rechner oder, allgemeiner, elektronische Informationsverarbeitungsma-schinen mit Eigenschaften, die sie für den massenhaften Einsatz in der Praxis geeignet machen. Maschinen mit solchen Eigenschaften können heute nur mit Hilfe der Mikroelektronik realisiert werden. Die Verhältnisse lassen sich in gewisser Hinsicht mit der Herstellung von Büchern vor und nach der Erfindung des Buchdrucks vergleichen. Sobald nur die Schrift erfunden war, bestand prinzipiell die Möglichkeit, beliebige Texte aufzuzeichnen und zu verbreiten. Dazu war es «nur» nötig, die richtigen Schriftzeichen in Steinplatten einzumeißeln. Wenn jeder Mensch auf diese Weise in zehnstündiger täglicher Arbeit *eine* «Buchseite» aus Stein beschrieben hätte, so wären vielleicht annähernd

so viele «Bücher» hergestellt worden wie benötigt wurden. Aber praktisch war das natürlich nicht durchführbar. Es waren technische Erfindungen nötig, die die massenhafte Herstellung von Büchern mit einem vertretbaren Aufwand an menschlicher Arbeitskraft praktisch ermöglichen. Die entscheidenden Schritte dabei waren erstens die Erfindung des Papiers und zweitens die des Buchdrucks. Mitunter wird die Mikroelektronik tatsächlich mit dem Buchdruck verglichen. Hier wird ihre Rolle eher unterschätzt als überschätzt. Der Buchdruck hat zwar einen sehr wichtigen, aber aus der Gesamtsicht doch auch sehr speziellen Informationsprozeß zum Gegenstand: die Speicherung von Information in Form von Buchstaben und Zeichen auf Papier. Demgegenüber ist die Mikroelektronik für den allgemeinsten Informationsprozeß, die Informationsverarbeitung, zuständig, und das in der ganzen Breite. Die Mikroelektronik bietet die erste umfassende technische Lösung für Informationsverarbeitungsprozesse in der bisherigen Entwicklung der menschlichen Gesellschaft. Es gibt nur noch *eine* Lösung, die der mikroelektronischen insgesamt gesehen überlegen ist, aber das ist keine technische, sondern eine natürliche Lösung, nämlich der denkende Mensch selbst. Mit Hilfe seines Gehirns verarbeitet der Mensch Information. Geistige Tätigkeit ist mit Informationsverarbeitung verbunden, wie umgekehrt Informationsverarbeitung eine geistige Leistung darstellt. Ist die Fähigkeit zur Informationsverarbeitung genügend entwickelt, so spricht man von Intelligenz — beim Menschen von

natürlicher, bei einer Maschine von künstlicher. Selbstverständlich wird damit nur eine *Analogie* zwischen Mensch und Technik zum Ausdruck gebracht, und keine Wesensgleichheit. Die Grenze, bis zu der die Informationsverarbeitung mit Hilfe der Mikroelektronik einfach «Informationsverarbeitung» ist und ab der sie in eine künstlich intelligente Leistung umschlägt, ist nicht sehr scharf definiert. Einig scheint man sich darüber zu sein, daß jenseits dieser Grenze die Fähigkeit zum Umgang mit *Wissen* vorhanden sein muß. Nicht selten wird eine Maschine aber auch dann bereits als «intelligent» bezeichnet, wenn sie *überhaupt* mit mikroelektronischer Informationsverarbeitungstechnik ausgestattet ist. Die Ungleichheit zwischen der künstlichen Intelligenz der Technik und der natürlichen des Menschen muß übrigens nicht unter allen Umständen und in jeder Hinsicht zu Ungunsten der Technik ausfallen. In zwei Punkten kann die Informationsverarbeitung mit Hilfe der Mikroelektronik der beim Menschen sogar überlegen sein, nämlich im Hinblick auf die Zuverlässigkeit und Schnelligkeit. Das menschliche Gehirn ist zwar in der Lage, außerordentlich komplexe Operationen auszuführen und wird darin von keiner Informationsverarbeitungsmaschine übertroffen. Aber der Mensch irrt sich relativ häufig, und das Gehirn arbeitet relativ langsam. Der Mensch kann in einer Minute etwa bis 100 zählen, ein Rechner auf der Basis der Mikroelektronik schafft es in der gleichen Zeit bis zu einer Million und darüber. Nicht schnell wie ein Gedanke, sondern langsam wie ein Gedanke müßte es hier eigentlich heißen. Aber nicht

diese teilweise Überlegenheit der mikroelektronischen Informationsverarbeitung über die beim Menschen ist das Entscheidende, obwohl auch daraus bedeutende Vorteile erwachsen. Entscheidend ist, daß in Gestalt der Mikroelektronik überhaupt eine praktikable technische Lösung für die Informationsverarbeitung existiert.

Welche Auswirkungen sich daraus für die Praxis konkret ergeben, soll im folgenden besprochen werden. Dabei ist zu beachten, daß die Mikroelektronik heute erst wenig über 20 Jahre alt ist, daß es den Mikroprozessor, der die breite Anwendung der Mikroelektronik vor allem gebracht hat, erst seit reichlich zehn Jahren gibt und daß sich auch heute in der Mikroelektronik noch bedeutende Entwicklungen vollziehen. Der Stand der Anwendung der Mikroelektronik für die Informationsverarbeitung, so wie er sich heute darstellt, ist deshalb alles andere als abschließend. Er ist ein Zwischenstadium, wenn auch ein überaus bedeutendes, in einem Entwicklungsprozeß, der noch wesentlich weitergetrieben werden kann und muß. Das, worüber im folgenden zu reden sein wird, sind deshalb neben den bereits erzielten Ergebnissen vor allem die Entwicklungstendenzen für die Anwendung der Mikroelektronik. Wir wollen uns mit vier dieser Tendenzen beschäftigen, die wir für besonders wichtig halten.

Besser als frühere Lösungen

Eine erste Tendenz hängt damit zusammen, daß zur Lösung von Problemen der Informationsverarbeitung technische Hilfsmittel angewandt worden sind, noch lange bevor es die Mikroelektronik gegeben

hat. Den Rechner auf der Basis von Elektronenröhren haben wir schon erwähnt. Aber auch elektromechanische oder sogar mechanische Rechner sind früher einmal benutzt worden. Elektromechanische Registrierkassen sind zum Teil heute noch im Einsatz. Ähnliches trifft auf elektromechanische Bedienungs- und Spielautomaten zu. Uhren für den Alltagsgebrauch funktionierten früher ausschließlich auf mechanischer oder elektrischer Grundlage. Anders als bei den Rechenmaschinen hat es elektronische Uhren mit Elektronenröhren oder einzelnen Halbleiterbauelementen, außer für wissenschaftliche Spezialzwecke, in der Praxis nie gegeben. Sie wären viel zu unhandlich und teuer gewesen. Bei Radio- und Fernsehgeräten wurden erst die Elektronenröhren und später die diskreten Halbleiterbauelemente verwendet. Heute werden bei den elektronischen Heimgeräten und in fast allen vorher genannten Fällen mikroelektronische Schaltkreise eingesetzt, und da, wo vielleicht doch noch andere Lösungen anzutreffen sind, wird der Übergang zur Mikroelektronik vorbereitet. Darin besteht eine erste wichtige, mit der Anwendung der Mikroelektronik verbundene Tendenz in der Entwicklung der Informationsverarbeitung. Diese Tendenz setzt sich deshalb durch bzw. hat sich bereits zum großen Teil durchgesetzt, weil die mikroelektronische Lösung wesentlich besser, d. h. billiger, zuverlässiger, schneller, platz-, material- und energiesparender ist als die vorhergehenden Lösungen. Für Informationsverarbeitungsprobleme, für die es über Jahrzehnte, bei der Uhr sogar über Jahrhunderte hinweg, andere tech-

nische Lösungen gegeben hat, bietet die Mikroelektronik heute eine wesentlich bessere. Am Beispiel der Rechenmaschine haben wir uns von der Richtigkeit dieser Aussage unmittelbar überzeugen können. Die Mikroelektronik ermöglicht den Bau kleiner und großer Rechner einer früher nicht gekannten Leistungsfähigkeit für einen relativ niedrigen Preis. Der Heimcomputer, der heute so viel leistet wie früher ein großer Rechner in einem Rechenzentrum und der trotzdem für den Einzelnen finanziell erschwinglich ist, bietet dafür ein Beispiel. Von der Leistungssteigerung bei großen Rechnern profitieren vor allem solche Tätigkeitsbereiche, die, wie die Wissenschaft, die Wettervorhersage, die Planung und Leitung von Produktions- und anderen gesellschaftlichen Prozessen, ohne Großrechner nicht auskommen. Was die Ablösung der Röhren und der diskreten Halbleiterbauelemente für die elektronischen Heimgeräte bedeutet, weiß jeder aus eigener unmittelbarer Erfahrung. Nicht nur daß diese Geräte kleiner, energiesparender, zuverlässiger und billiger wurden — durch den Einsatz der Mikroelektronik sind qualitativ neue, in ihrem Gebrauchswert wesentlich verbesserte Geräte entstanden, die sich durch höhere Ton- und Bildqualität, durch höheren Bedienungskomfort und durch zusätzliche Funktionen auszeichnen.

Technische Lösungen für früher unlösbare Probleme. Automatisierung
Für die meisten Probleme der Informationsverarbeitung gab es aber vor der Erfindung der Mikroelektronik überhaupt keine technische Lösung. Die Mikroelektronik macht

in vielen dieser Fälle erstmalig eine solche Lösung möglich. Mit Hilfe der Mikroelektronik können heute und in Zukunft Informationsverarbeitungsprobleme gelöst werden, die früher entweder überhaupt nicht oder nur durch den Menschen selbst gelöst werden konnten. Darin besteht eine zweite wichtige, auf der Anwendung der Mikroelektronik beruhende Entwicklungstendenz der Informationsverarbeitungstechnik. Die Informationsverarbeitung in der materiellen Produktion war früher, soweit sie überhaupt erfolgte, weitgehend die Sache des Menschen selbst. Seitdem leistungsfähige elektronische Großrechner zur Verfügung stehen, war es im Prinzip möglich, die in der Produktion anfallenden Probleme der Informationsverarbeitung maschinell zu lösen. Dazu war es aber nötig, die Information von den verschiedenen Punkten des Produktionsprozesses an den zentral gelegenen Großrechner weiterzuleiten, sie dort zu verarbeiten und danach wieder zurückzuleiten. Aber der Großrechner war viel zu teuer und das Verfahren insgesamt viel zu aufwendig und störanfällig, um der elektronischen Informationsverarbeitung in der Produktion auf breiterer Front zum Durchbruch zu verhelfen. Den entscheidenden Fortschritt hat auch hier erst die Mikroelektronik gebracht. Durch die Mikroprozessoren und Mikrorechner, die seit Anfang der siebziger Jahre für einen relativ niedrigen Preis zur Verfügung stehen, ist es möglich geworden, jede einzelne Maschine oder Anlage mit einem eigenen kleinen Zentrum für die Informationsverarbeitung auszustatten. Die Information kann auf diese

Weise unmittelbar da, wo sie entsteht bzw. benötigt wird, verarbeitet werden. Die Informationsverarbeitung wird dezentralisiert. Der Einsatz kleiner Rechner vor Ort statt des entfernt gelegenen Großrechners ist für die Informationsverarbeitung das, was der Übergang von der Dampfmaschine zum Elektromotor für die Antriebsversorgung von Maschinen war. Mit einer Dampfmaschine als Antriebsaggregat konnte in der Regel nicht jede einzelne Maschine, sondern nur ein ganzer Komplex mehr oder weniger weit voneinander entfernt gelegener Maschinen ausgerüstet werden. Mittels des Elektromotors kann dagegen jede Maschine ihr eigenes Antriebsaggregat erhalten, ähnlich wie mittels des Mikrorechners ihr eigenes Aggregat für die Informationsverarbeitung. Für die Steuerung derart ausgestatteter Maschinen wird der Mensch nicht mehr benötigt. Sie steuern sich mit Hilfe des Mikrorechners und des in ihm gespeicherten Programms selbst. Man spricht deshalb von automatischer Steuerung. Mit automatisch gesteuerten Maschinen und Anlagen lassen sich ganze Produktionsprozesse automatisieren. Je nach dem Grad und der Art der Automatisierung unterscheidet man dabei verschiedene Formen. Das Ziel ist zweifellos der vollständig automatisierte Produktionsprozeß, in dem der Mensch nur noch eine mittelbare Funktion ausübt. Am ehesten ist dieses Ziel heute bei der Herstellung von Produkten erreicht, die in exakt der gleichen Form über einen längeren Zeitraum in großer Menge benötigt werden, wie z. B. Eisenbahnschienen, elektrische Kabel, Zement oder chemische Grund-

stoffe. Die Herstellung dieser und ähnlicher Produkte erfolgt auf Fertigungsstraßen oder in Anlagen, die, ebenso wie die Produkte selbst, über einen längeren Zeitraum hinweg nicht verändert zu werden brauchen. Die Automatisierung derartiger Einzweck-Fertigungssysteme mit Hilfe mikroelektronischer Steuerungstechnik wird heute gut beherrscht und ist von hohem ökonomischen Nutzen. Man spricht dabei von Einzweck-Automatisierung. Aber die wenigsten Produkte sind so dauerhaft und können unverändert über viele Jahre hinweg in so großem Umfang produziert werden wie Eisenbahnschienen und Zement. In vielen Fällen sind nur kleine oder mittlere Serien eines bestimmten Produktes gefragt, und durch technische Neuerungen oder aus Modegründen werden innerhalb relativ kurzer Zeitabstände Veränderungen an den Produkten erforderlich. Die Einzweck-Automatisierung ist in diesen Fällen technisch zwar ebenso gut möglich wie bei Massenprodukten, ökonomisch ist sie aber in der Regel unvorteilhaft. Es ist eine andere Automatisierungslösung erforderlich, um einen ähnlich hohen ökonomischen Nutzen erzielen zu können, wie ihn die Einzweck-Automatisierung bei der Herstellung dauerhafter Massenprodukte möglich gemacht hat. Ansatzpunkte zu einer solchen Lösung waren die numerisch gesteuerten Werkzeugmaschinen und die Industrieroboter, wie sie sich bereits Ende der 50er, Anfang der 60er Jahre entwickelt haben. Durch Kombination von Maschinen, deren Abläufe gesteuert und deren Werkzeuge ausgewechselt werden können, mit Robotern für die

Beschickung der Maschinen und den Transport der Werkstücke, entstehen sog. flexible Fertigungssysteme. Die Flexibilität dieser Systeme besteht darin, daß sie durch relativ geringfügige Änderungen — wie eben etwa den Werkzeugwechsel — auf die Herstellung unterschiedlicher Produktformen eines bestimmten Grundtyps umgestellt werden können. Um ein verändertes Produkt herstellen zu können, ist nicht ein neues Fertigungssystem erforderlich, sondern das vorhandene System braucht nur an das neue Produkt angepaßt zu werden. Dadurch werden mit geringen zusätzlichen Aufwendungen für die Anpassung an die konkrete Produktform hohe Stückzahlen eines bestimmten Produkt-Grundtyps und kurze Innovationszeiten möglich. Um etwas Konkretes vor Augen zu haben, können wir uns den Produkt-Grundtyp als Maschinenteil, Glühlampe oder Verpackungsbehälter vorstellen und die konkrete Produktform als Welle, 60-Watt-Lampe mit Normalfassung oder als Waschmittelpackung. In Verbindung mit den flexiblen Fertigungssystemen ermöglicht die Mikroelektronik die gesuchte neue Lösung für die Produktionsautomatisierung, die sogenannte *flexible Automatisierung*. Ziel der flexiblen Automatisierung ist die rechnergestützte Fertigung eines schnell wechselnden Sortiments von Produkten in kleiner und mittlerer Stückzahl mit den gleichen Produktionsmitteln (CAM — *Computer Aided Manufacturing*), die rechnergestützte Konstruktion dieser Produkte am Bildschirmarbeitsplatz (CAD — *Computer Aided Design*) und möglichst die Verknüpfung zwischen

beiden, ja die Integration des gesamten Produktionsprozesses, angefangen von der Konstruktion, der Materialbereitstellung in Verbindung mit der Lagerhaltung und dem innerbetrieblichen Transport über die eigentliche Fertigung, die Qualitätskontrolle und die Verpackung bis hin zur Auslieferung des Produktes, zu einem einzigen rechnergestützten Gesamtprozeß. Diese Integration fällt ökonomisch um so mehr ins Gewicht, als gerade im Umfeld des eigentlichen Herstellungsprozesses große Reserven für die Einsparung von Arbeitskräften und für die Verkürzung der Durchlaufzeiten der Produkte verbunden mit einer höheren Auslastung der Maschinen liegen. In vielen Ländern, darunter auch in der DDR, wird heute intensiv an der Schaffung der Voraussetzungen für die flexible Automatisierung gearbeitet. Zu diesen Voraussetzungen gehören insbesondere die Industrieroboter. Wie der Name sagt, führen Industrieroboter Tätigkeiten im Produktionsprozeß aus, die sonst vom Menschen verrichtet werden müssen. Industrieroboter werden heute bereits an vielen Stellen in der Produktion, auch außerhalb flexibler Fertigungssysteme, für die unterschiedlichsten Arbeiten eingesetzt, so z. B. beim Lackieren, Schweißen oder Beschicken von Maschinen. Im Rahmen automatisierter flexibler Fertigungssysteme können sie ihre Wirksamkeit voll entfalten. Man unterscheidet Industrieroboter nach Generationen. Die Industrieroboter der ersten Generation sind automatisch gesteuerte Manipulatoren, wobei die Steuerung von der ein für allemal fixierten mechanischen bis hin zur frei programmierbaren mikro-

elektronischen reichen kann. Roboter der zweiten und dritten Generation besitzen außer der mikroelektronischen Steuerung noch Sensoren, mit deren Hilfe sie den Zustand des Arbeitsgegenstandes und der Werkzeuge erfassen können und — unter Nutzung dieser Information — das Arbeitsprogramm der aktuellen Situation anpassen können. Charakteristisch für die Roboter der dritten Generation ist ihr entwickeltes Vermögen für die Bilderkennung und zur darauf abgestimmten Koordinierung der Bewegungsabläufe. Roboter der vierten Generation besitzen künstliche Intelligenz, d. h. eine höhere Fähigkeit zur Informationsverarbeitung als ihre Vorgänger. Die zur Zeit in der Industrie eingesetzten Roboter gehören im wesentlichen noch zur ersten Generation.

Die Automatisierung der Arbeit schreitet heute nicht nur in der Produktion, sondern auch in der Verwaltung mit großen Schritten voran. Auch hier hatte die zentrale elektronische Datenverarbeitung mit Hilfe großer Rechner zunächst nicht den erhofften Erfolg gebracht, und der Durchbruch erfolgte ebenfalls erst durch den Übergang zur dezentralen Informationsverarbeitung im einzelnen Büro bzw. an der einzelnen Maschine mit Hilfe der Mikroelektronik. Ähnlich wie intelligente Werkzeugmaschinen in der Produktion sind heute in den Büros intelligente Schreibmaschinen im Einsatz. Die umfassendste und universellste Lösung für die Automatisierung von Büroprozessen ist gegenwärtig der Bürocomputer. Mit seiner Hilfe kann das Speichern, Verarbeiten und Übertragen von Information als rechnergestützter Gesamtprozeß

durchgeführt werden. Das Papier als Anzeige- und Speichermedium wird dabei in zunehmendem Maße durch den Bildschirm bzw. durch magnetische Datenträger wie Disketten oder Magnetbänder ersetzt. Für die Zukunft wird das «papierlose» Büro vorausgesagt.

Die Automatisierung von Prozeßabläufen durch Informationsverarbeitung mit Hilfe der Mikroelektronik ist aber nicht auf die materielle Produktion und Verwaltung beschränkt. Im Haushalt können durch den Einsatz der Mikroelektronik für eine Reihe von Prozessen wie das Wäschewaschen, Kochen oder Geschirrspülen voll- oder teilautomatische Maschinen zur Verfügung gestellt werden. In der naturwissenschaftlichen und technischen Forschung sind es die *Meßprozesse*, die mit Hilfe der Mikroelektronik automatisiert werden können. Die hohe Geschwindigkeit, mit der die mikroelektronische Informationsverarbeitung vonstatten geht, ermöglicht auch die automatische Steuerung sehr schnell ablaufender Prozesse, wie sie etwa das Landen eines Flugzeuges oder der Flug einer Rakete darstellen. Bei anderen Prozessen kann man es sich leisten, umfangreiche Meßinformationen über den Prozeßablauf aufzunehmen und aufwendigen Rechenoperationen zu unterwerfen, um auf diese Weise die Prozeßführung unter bestimmten Gesichtspunkten zu optimieren. Bei energieintensiven Prozessen kann ein solcher Gesichtspunkt z. B. die Minimierung des Energieverbrauchs sein. Die Mikroelektronik ermöglicht also nicht nur die Automatisierung, sondern auch die *Optimierung von Prozeßabläufen*.

Wenn von technischen Lösungen für die Informationsverarbeitung die Rede ist, die erst durch die Mikroelektronik möglich geworden sind, dürfen auch die elektronischen Spielgeräte nicht fehlen. Spielautomaten gab es auch schon früher, aber ihre Intelligenz war, sofern überhaupt vorhanden, sehr gering. Der erste Schachautomat — ein überdimensionierter «Türke», der in der zweiten Hälfte des 18. Jahrhunderts in Europa und Amerika großes Aufsehen erregte — beruhte auf einem Betrugsmanöver. In der Maschine war ein Mensch versteckt, der sehr gut Schach spielen konnte. Heutige Schachcomputer kommen ohne diesen Trick aus. Sie besitzen eine künstliche Intelligenz in Form von mikroelektronischen Schaltkreisen, die ausreicht, um auch einen geübten Schachspieler in Schwierigkeiten zu bringen. Die vielen anderen elektronischen Spiele, die heute auf dem Markt sind, funktionieren nach dem gleichen Prinzip. Die intelligente Mikroelektronik des Gerätes übernimmt die Rolle des Gegenspielers. Unterstützt wird sie bei den Videospielen u. U. noch von einem Programm, das man über eine Videokassette in das Gerät eingibt, und einen Fernsehapparat, auf dessen Bildschirm das Spielgeschehen sichtbar wird.

Auswirkungen

auf andere Informationsprozesse Neue Kommunikationstechnologien

Bisher war ausschließlich von der Bedeutung der Mikroelektronik für die Informationsverarbeitung die Rede. Wir kommen jetzt auf Auswirkungen der Mikroelektronik auf die anderen Informationsprozesse zu sprechen. Diese Auswirkungen be-

ruhen darauf, daß mehr oder weniger alle Informationsprozesse, also auch die Gewinnung, Speicherung oder Übertragung von Information, unmittelbar mit Verarbeitungsschritten verbunden sind. Mit Hilfe der Mikroelektronik können auch diese Verarbeitungsschritte auf einem höheren Niveau durchgeführt werden als früher. Dadurch erhöht sich insgesamt das Niveau der technischen Lösungen für diese Informationsprozesse. Darin besteht eine dritte wesentliche Auswirkung der Mikroelektronik auf die Informationstechnik. Sie findet z. B. ihren konkreten Ausdruck darin, daß Speicher und vor allem Sensoren für die Informationsgewinnung mit Hilfe der Mikroelektronik intelligent gemacht, d. h. mit eigenen Informationsverarbeitungseinheiten ausgerüstet werden können. Auch die umwälzenden Entwicklungen, die sich heute, für jedermann sichtbar, auf dem Gebiet der Telekommunikation vollziehen, haben die mikroelektronische Lösung für die Informationsverarbeitung zur Voraussetzung. Freilich lassen sich diese Entwicklungen nicht auf die Anwendung der Mikroelektronik reduzieren, sondern beruhen auch noch auf anderen technischen Neuerungen wie dem Einsatz von Nachrichtensatelliten, der Verlegung von zusätzlichen Kupferkabeln und dem Übergang zu Glasfaserkabeln für die Lichtleiterübertragung. Die zuletzt genannten Verbesserungen betreffen die Übertragungswege, die sogenannten *Informationskanäle*. Es werden mehr Kanäle mit größerer Übertragungskapazität zur Verfügung gestellt. Über sie können mehr Rundfunk- und Fernsehprogramme und ein

breites Spektrum anderer Informationen als Bildschirmtext übertragen werden. Der elektronische Brief- und Zahlungsverkehr, die elektronische Zeitung und das Bildtelefon werden möglich. Realisierbar erscheint auch die Kommunikation einzelner Teilnehmer im Kabelnetz mit den großen Informationsspeichern, wie z. B. elektronischen Datenbanken und Bibliotheken, oder mit den großen elektronischen Rechnern. Durch die neuen Kommunikationstechnologien könnte es dahin kommen, daß der einzelne Teilnehmer über diese Informationsspeicher und -verarbeitungsmaschinen so verfügen kann, als ob er sie bei sich zu Hause in der Wohnung hätte. Statt in die Bibliothek zu gehen, das Buch auszuleihen und mit nach Hause zu nehmen, kann er sich dann in seiner Wohnung vor den Fernsehapparat setzen und die Angaben über das gewünschte Buch in einen Heimcomputer eingeben, der mit dem Telefonkabel verbunden ist. Nach einigen Sekunden erscheint der Buchtext auf dem Bildschirm. Statt in ein Reisebüro zu gehen und zu erfragen, welche Reisen angeboten werden und ob dafür noch Plätze zu haben sind und schließlich eine bestimmte Reise durch Ausfüllen von Formularen, Unterschriftsleistung und Bezahlung zu buchen, kann er alles einfach mit Hilfe des an das Telefonkabel angeschlossenen Heimcomputers und des Fernsehapparates erledigen. Was hier an Beispielen aus dem privaten Bereich beschrieben wurde, ist auf gesellschaftlicher Ebene im Bereich der materiellen Produktion, in der Verwaltung oder bei der Planung und Leitung gesellschaftlicher Prozesse noch von weit größerer Bedeutung.

Der gesamte dienstliche Briefverkehr kann auf elektronischem Wege durchgeführt werden. Informationen, die für Entscheidungen benötigt werden, können in Sekundenschnelle zur Verfügung gestellt und verarbeitet werden. Beratungen, zu denen viele Teilnehmer über Hunderte von Kilometern hinweg an einen bestimmten Ort reisen müssen, um sich dort zu treffen, lassen sich mit Hilfe der neuen Kommunikationstechnologien so durchführen, daß sich keiner der Teilnehmer aus seinem Arbeitsraum zu entfernen braucht. Jeder ist auf dem Bildschirm für den anderen sichtbar und durch den Lautsprecher hörbar. Daß im Zahlungsverkehr statt der Beförderung von Papiergeld oder von Münzen genausogut auch Informationen auf elektronischem Wege ausgetauscht werden können, hatten wir bereits erwähnt. Den angeführten Beispielen ist eines gemeinsam: es handelt sich um Fälle der Übertragung von Information, die heute weitgehend noch so gelöst werden, daß beträchtliche Mengen stofflicher Materie in Bewegung gesetzt werden, statt wie bei der elektronischen Übertragung nur ganz winzige Mengen von Energie. Die Information wird dabei transportiert, indem der materielle Träger, auf dem sie gespeichert ist, transportiert wird. Der Übergang zur elektronischen Informationsübertragung bedeutet in diesen Fällen, daß Material- und Energieströme durch Informationsströme ersetzt werden. Diese Formulierung stellt zwar eine Übertreibung dar, aber sie trifft das Wesentliche. Informationsströme sind natürlich ohne Energieströme genausowenig denkbar wie die Farbe ohne den Stoff. Aber die

Energieströme können bei der elektronischen Informationsübertragung außerordentlich klein gemacht werden, so daß sie gegenüber den Informationsströmen, die sie übertragen, völlig in den Hintergrund treten. Diese sich abzeichnenden Fortschritte auf dem Gebiet der Telekommunikation sind ebenso wie die heute bereits realisierten an die Bereitstellung von mehr und leistungsfähigeren Übertragungskanälen gebunden: Das ist aber nur die eine Seite. Um Information durch einen Kanal übertragen zu können, muß sie vorher vom primären Informationsträger abgenommen, in den Übertragungskanal eingespeist und nach der Übertragung aus ihm wieder entnommen und auf dem Bildschirm oder anderweitig dargestellt werden, d. h., die Information muß verarbeitet werden. Um eine sichere Wiedererkennung der übertragenen Information gewährleisten zu können, müssen Codierungsverfahren angewandt werden, die einen zusätzlichen Verarbeitungsaufwand mit sich bringen. Soll viel Information pro Zeiteinheit übertragen werden, muß auch viel Information pro Zeiteinheit verarbeitet werden. Die modernen Kommunikationsverfahren erfordern also eine schnelle Verarbeitung großer Informationsmengen. Die einzig praktikable Lösung für dieses Problem bietet die Mikroelektronik. Ohne sie sind die modernen Kommunikationstechnologien nicht realisierbar. Durch die Mikroelektronik können die Informationsübertragung und die anderen Informationsprozesse auf ein wesentlich höheres Niveau als früher gehoben werden. Darin eben besteht ihre dritte wichtige Auswirkung.

Einheitliches

Informationssystem

Diese Auswirkung hat aber selbst wieder Konsequenzen. Wenn Informationsverarbeitung mit Hilfe der Mikroelektronik zu einem wesentlichen Bestandteil auch der übrigen Informationsprozesse wird, so werden dadurch die Grenzen zwischen den verschiedenen Informationsprozessen fließend. Der gleiche Effekt tritt ein, wenn dank der neuen Kommunikationstechnologien große Informationsmengen zwischen entfernt voneinander gelegenen Informationsverarbeitungszentren und großen Informationsspeichern ausgetauscht werden können. Dadurch rücken die Verarbeitungszentren und die Speicher eng zusammen. Sie werden zu einem einzigen System, in das auch die Übertragungseinheiten mit eingeschlossen sind. In diesem System lassen sich Informationsverarbeitungs-, -speicherungs- und -übertragungsprozesse überhaupt nicht mehr voneinander trennen. Die Lösung eines konkreten Informationsproblems löst gleichzeitig alle drei Prozesse aus. Es erfolgt eine Verschmelzung von Informationsverarbeitung, -speicherung und -übertragung. Im Bereich der Energiewirtschaft ist diese Entwicklung mit der Herausbildung der landesweiten und regionalen Verbundsysteme vergleichbar. Die Integration der Teilsysteme für die einzelnen Informationsprozesse zu einem einzigen System für den Gesamtinformationsprozeß ist heute eine klar erkennbare vierte Tendenz in der Entwicklung der Informationstechnik. Auch sie beruht wesentlich auf dem Einsatz der Mikroelektronik.

Mikroelektronik und wissenschaftlich-technische Revolution

Nach der Behandlung dieser und anderer Einzelpunkte zum Verhältnis zwischen Mikroelektronik, Information und Gesellschaft wollen wir uns zum Schluß noch einmal dem allgemeinen Anliegen des vorliegenden Kapitels zuwenden, nämlich herauszufinden, worin die Bedeutung der Mikroelektronik für die Gesellschaft eigentlich besteht und warum an sie so hohe Erwartungen geknüpft werden. Wie wir gesehen haben, gibt es dafür mehrere Gründe. Der entscheidende Grund ist, daß die Mikroelektronik nicht nur zu dieser oder jener technischen Neuerung führt, sondern daß sie die Technik von Grund auf revolutioniert. Die Technik, wie sie im Ergebnis der industriellen Revolution im 18. und 19. Jahrhundert geschaffen wurde, diente im wesentlichen der Verrichtung körperlicher Arbeit. Durch die Mikroelektronik wird eine Technik möglich, die geistige Arbeit oder neben physischer *auch* geistige Arbeit verrichten kann. Mit Recht wird die Schaffung einer solchen Technik als Hauptinhalt der wissenschaftlich-technischen Revolution unserer Tage angesehen und mit der industriellen Revolution früherer Tage auf eine Stufe gestellt. Ähnlich wie die industrielle Revolution den Menschen von der Last schwerer körperlicher Arbeit befreite und ihm Werkzeuge in die Hand gab, mit deren Hilfe er seine physischen Kräfte ins nahezu Unbegrenzte steigern konnte, befreit ihn die wissenschaftlich-technische Revolution heute zunehmend von der Last routinemäßiger geistiger Arbeit und versetzt ihn durch die

Bereitstellung von «Denk-Werkzeugen» in die Lage, seine intellektuellen Kräfte in einem früher nicht vorstellbaren Maße zu vergrößern. Informationsverarbeitende Prozesse, die der Mensch bisher selbst durchführen mußte, nimmt ihm die Technik ab. Gewohnte Arbeitsaufgaben fallen weg, neue entstehen und mit ihnen neue Qualifikationsanforderungen und Tätigkeitsmerkmale für viele Werkstätige. Im Extremfall können bestimmte traditionelle Berufe überhaupt entbehrlich werden. Derartige Vorgänge sind Begleiterscheinungen der Entwicklung der Produktivkräfte durch die Anwendung der Mikroelektronik im Zuge der wissenschaftlich-technischen Revolution. Diese Entwicklung führt zu einer gewaltigen Steigerung der Produktivität gesellschaftlicher Arbeit und ermöglicht eine beträchtliche Vermehrung des materiellen Reichtums der Gesellschaft. Aber diese Entwicklung kostet auch ihren Preis. Es ist nur recht und billig, wenn dieser Preis von der Gesellschaft als Ganzes entrichtet wird, und wenn an der Vermehrung des Reichtums alle gleichermaßen Anteil haben. In der Entwicklung der Produktivkräfte mußte die Mikroelektronik früher oder später kommen. Sie mußte erfunden werden, weil das gesellschaftliche Bedürfnis, dem sie Abhilfe schaffen sollte, die Übertragung geistiger Arbeit an die Technik, vorhanden war. Unter diesem Gesichtspunkt ist es eigentlich ganz zufällig, daß eine Mikroelektronik entstanden ist, die auf der Basis von Halbleitereffekten arbeitet. Es hätten auch andere Wirkprinzipien — physikalische, chemische oder biologische — genommen werden können,

entscheidend war nur, daß sich mit ihrer Hilfe Informationsverarbeitungsprobleme technisch so lösen lassen, wie es die gesellschaftliche Praxis erfordert. Daß die Mikroelektronik heute so ist wie sie ist, daß sie Halbleiterphysikalische Effekte als Wirkprinzipien und Silizium als Materialbasis verwendet, daß es nicht mehrere Mikroelektroniken nebeneinander gibt, ist aus der Sicht der naturwissenschaftlich-technischen Entwicklung allerdings alles andere als zufällig. Darauf werden wir im folgenden Kapitel genauer eingehen.

3. Halbleiterbauelemente

Historisch gesehen hat sich die Elektronik zuerst der Elektronen im Vakuum bedient. Man denke etwa an eine Elektronenröhre, bei der sich die Elektronen in einem evakuierten Glaskolben bewegen. Später, vor allem seit der Erfindung des Transistors im Jahre 1948 durch Bardeen, Brattain und Shockley, sind dazu die Elektronen in Halbleiterkristallen gekommen und haben die Elektronen im Vakuum zurückgedrängt, ohne sie jedoch ganz zu ersetzen. Die Mikroelektronik verwendet ausschließlich Elektronen in Halbleitern.

Halbleiter

Halbleiter sind Kristalle aus chemischen Elementen, wie z. B. Silizium, Germanium, Selen oder Tellur, und aus chemischen Verbindungen, wie z. B. Galliumarsenid, Cadmiumsulfid oder Bleitellurid. Der Begriff «Halbleiter» bedeutet zunächst nicht mehr, als daß es sich dabei um einen elektrischen Leiter handelt, dessen spezifische Leitfähigkeit zwischen der von Isolatoren ($< 10^{-10} \Omega^{-1} \text{ cm}^{-1}$) und der von Metallen ($> 10^2 \Omega^{-1} \text{ cm}^{-1}$) liegt. Diese Charakterisierung eines Halbleiters ist relativ äußerlich. Ihr liegen aber wesentliche innere Besonderheiten zugrunde, über die im folgenden noch zu sprechen sein wird und die den eigentlichen Wert der Halbleiter für die Elektronik ausmachen. Der Schritt der Elektronik vom Va-

kuum zum Halbleiter ist logisch. Wenn es der Elektronik nur darauf ankommt, frei bewegliche Elektronen zur Verfügung zu haben und wenn sich solche auch in festen Körpern vorfinden, so ist es überflüssig, diese Elektronen dadurch zu erzeugen, daß man sie aus dem festen Körper erst ins Vakuum austreten läßt, wie das bei der Elektronenröhre geschieht. Man kann gleich die Elektronen im Festkörper verwenden. Festkörper mit frei beweglichen Elektronen sind Metalle und Halbleiter. Warum nutzt die Elektronik ausschließlich Halbleiter und nicht auch Metalle als Grundmaterialien für ihre Bauelemente? Metalle und Halbleiter unterscheiden sich durch die Größe ihres Gehaltes an frei beweglichen Elektronen. In Metallen ist dieser Gehalt sehr groß, auf ein Atom des Metalls kommt etwa ein frei bewegliches Elektron. Die Dichte beträgt also ungefähr 10^{22} pro Kubikzentimeter. In Halbleitern gibt es dagegen relativ wenig frei bewegliche Elektronen, erst auf jedes 10^{10} -te bis 10^3 -te Atom kommt ein freies Elektron. Die Dichte liegt also zwischen 10^{12} und 10^{19} pro Kubikzentimeter. Wie wir wissen, besteht das Ziel der Elektronik darin, die Bewegung der Elektronen durch äußere elektrische Einwirkungen zu beeinflussen. Daß eine solche Beeinflussung bei den relativ wenigen

Elektronen eines Halbleiters leichter möglich sein wird als bei den relativ vielen eines Metalls, ist aber ohne weiteres verständlich. Für elektronische Bauelemente sind also Halbleiter besser geeignet als Metalle. Um zu verstehen, wie Halbleiterbauelemente funktionieren, müssen wir uns mit der Bewegung der Elektronen in einem Halbleiterkristall etwas genauer beschäftigen.

Energiebänder

Wir beginnen mit der Energie, die ein Elektron in einem solchen Kristall besitzen kann. Bei den Elektronen eines freien Atoms sind bekanntlich nicht beliebige Energiewerte möglich, sondern nur ganz bestimmte Einzelwerte, wie sie durch das Bohrsche Atommodell bzw. die Gesetze der Quantenphysik vorgegeben sind. Zwischen den erlaubten Einzelwerten liegen verbotene Energiebereiche, wo sich kein Elektron aufhalten kann. Ähnlich ist es bei einem Halbleiterkristall, nur daß die einzelnen erlaubten Energiewerte beim Zusammenbau der Atome zum Kristall in Energiebänder auffächern (Abb. 4). Zwischen den erlaubten Energiebändern liegen wie beim Atom Lücken ohne erlaubte Energiewerte. Die Elektronen des Kristalls müssen Energiewerte annehmen, die innerhalb der Energiebänder liegen. Jedes Energieband hat aber nur für eine bestimmte Anzahl von Elektro-

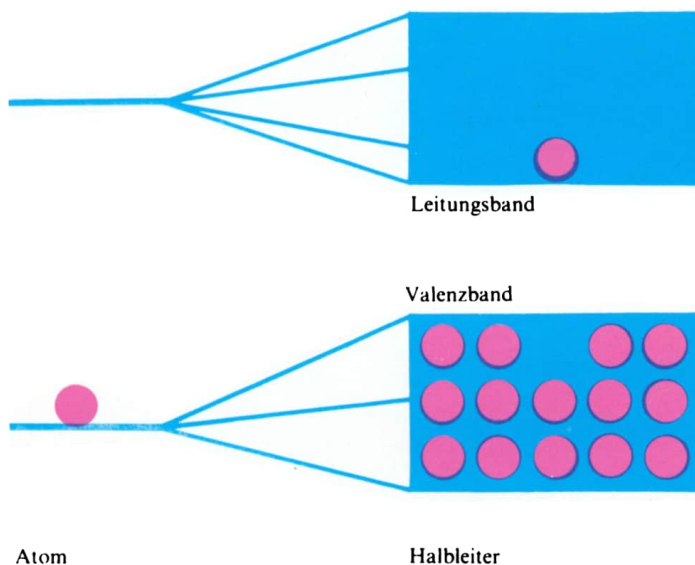


Abb. 4 Die Energieniveaus eines Atoms und die Energiebänder eines Halbleiters. Die schrägen Striche deuten an, wie sich die diskreten Energieniveaus beim Zusammenbau der Atome zum Kristall zu Energiebändern verbreitern.

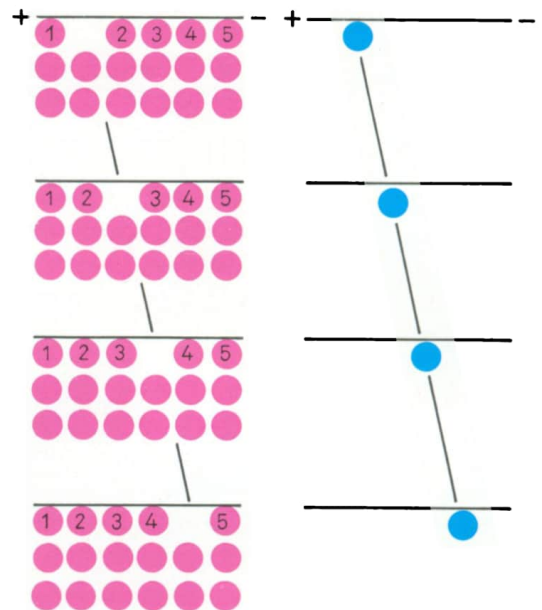


Abb. 5 Ein Loch in einem sonst vollständig mit Elektronen besetzten Band bewegt sich bei Vorhandensein einer elektrischen Spannung so wie ein positiv geladenes Teilchen.

nen Platz. Sind alle Plätze besetzt, läuft das Band über, und das Band mit der nächsthöheren Energie muß aufgefüllt werden. Halbleiter zeichnen sich dadurch aus, daß ein Energieband — das *Valenzband* — fast vollständig besetzt ist und das darüberliegende Energieband — das *Leitungsband* — fast vollständig leer ist. «Fast vollständig» heißt hier, daß bei 10^{22} Plätzen nur jeder 10^3 -te bis 10^{10} -te Platz leer bzw. besetzt ist. Die wenigen nichtbesetzten Plätze des Valenzbandes, die *Löcher*, verhalten sich wie positiv geladene Elektronen. Das ist leicht einzusehen. Wir legen dazu an den Halbleiter eine Spannung an. Was passiert mit

den Elektronen im Valenzband? Sie werden durch das elektrische Feld nach links gezogen (siehe Abb. 5). Sie können aber nur nach links rücken, wenn sie vor sich einen freien Platz haben. Das ist in Abb. 5 oben nur bei dem mit (2) bezeichneten Elektron der Fall. Dieses Elektron rückt also nach links in die Lücke, die Lücke rückt um eine Stelle nach rechts. Danach kann das Elektron (3) nach links in die Lücke rücken, die Lücke um einen weiteren Platz nach rechts. Die Lücke bewegt sich auf die negative Elektrode zu, d. h., sie bewegt sich so, wie sich ein positiv geladenes, frei bewegliches Teilchen bewegen würde. Die Löcher im Valenzband kann man also ebenso wie die Elektronen im Leitungsband als frei bewegliche Ladungs-

träger auffassen. Im Unterschied zu den Elektronen sind es aber positive Ladungsträger. Das Auftreten von zwei beweglichen Ladungsträgersorten, den negativen Elektronen und den positiven Löchern, ist eine der wichtigsten Besonderheiten von Halbleitern. Sie bildet zusammen mit der leichten Beeinflussbarkeit der Ladungsträgerkonzentration durch äußere Einwirkungen, auf die wir gleich noch zu sprechen kommen werden, die Grundlage für die Wirkungsweise von Halbleiterbauelementen. Halten wir uns noch einen Moment bei dem etwas merkwürdigen freien positiven Ladungsträger «Loch» auf. Ein Loch besteht letzten Endes auch aus Elektronen — es ist die Gesamtheit aller Elektronen des Va-

lenzbandes bis auf eines. Statt der Bewegung des Loches könnte man auch die Bewegung der Elektronen des bis auf einen Platz voll gefüllten Valenzbandes ins Auge fassen. Man müßte dann angeben, so wie wir es oben bei der Erläuterung von Abb. 5 getan haben, wie sich jedes der 10^{22} -1 Elektronen des Valenzbandes verschiebt. Das ist sehr aufwendig und zudem in hohem Maße überflüssig, denn die Bewegungsmöglichkeiten der Elektronen sind weitgehend festgelegt. Nur die Elektronen 1 und 2 im oberen Teil von Abb. 5 können sich verschieben, dabei 1 nur nach rechts und 2 nur nach links. Allen übrigen Elektronen ist eine Verschiebung verboten. Die Einführung der Löcher als frei bewegliche Ladungsträger im Valenzband ermöglicht also eine weit zweckmäßigere Beschreibung der Bewegung der Elektronen im Valenzband. Den Begriff des Loches könnte man im Prinzip auch auf das Leitungsband anwenden. Da das Leitungsband nur wenige Elektronen enthält, müßte man dann sagen, daß das Leitungsband fast vollständig mit Löchern gefüllt ist. Die Löcherbeschreibung ist beim Leitungsband aber ebenso unzweckmäßig wie die Elektronenbeschreibung beim Valenzband. Halbleiter mit vorwiegend Elektronen nennt man *n(negativ)-leitend*, solche mit vorwiegend Löchern *p(positiv)-leitend*. Die Einschränkung «vorwiegend» ist deshalb nötig, weil es in einem n-Halbleiter immer auch einige Löcher und in einem p-Halbleiter immer auch einige Elektronen gibt. Man nennt sie *Minoritätsladungsträger* im Unterschied zu den hauptsächlich anzutreffenden *Majoritätsladungsträgern*.

Um zu demonstrieren, daß es sich bei den Löchern nicht um etwas handelt, was nur in der Phantasie der Physiker existiert, wollen wir in Gedanken ein Experiment machen, das leicht auch in Wirklichkeit ausgeführt werden kann und in Abb. 6 schematisch dargestellt ist. Wie läßt sich entscheiden, ob in einer Halbleiterprobe hauptsächlich Elektronen oder Löcher enthalten sind? Offensichtlich nicht einfach durch die Richtung des Stromes, der unter der Wirkung einer bestimmten Spannung fließt. Zwar werden Elektronen und Löcher durch eine elektrische Spannung in entgegengesetzter Richtung bewegt, aber durch die entgegengesetzte Ladung, die beide tragen, hat ein nach rechts gerichteter Elektronenstrom die gleiche elektrische Stromrichtung wie ein nach links gerichteter Löcherstrom. Ein Gleichstrommesser stellt also keinen Unterschied fest. Ein Unterschied ergibt sich erst, wenn ein Magnetfeld senkrecht zur Stromrichtung angelegt wird. Das Magnetfeld krümmt die Bahn der Elektronen und Löcher, und zwar in einer Weise, die nur von der elektrischen Stromrichtung abhängt (Hall-Effekt), also in

p- und n-Halbleitern nach der gleichen Seite. Damit entsteht aber im p-Halbleiter ein nach unten gerichteter Strom, im n-Halbleiter ein nach oben gerichteter Strom. Die Existenz der Löcher kann auf diese Weise für jedermann sichtbar gemacht werden.

Dotierung

Wir haben bisher angenommen, daß im Halbleiter ohne unser Zutun p- oder n-Leitung vorhanden ist. Um die n- oder p-Leitung bewußt einzustellen, muß man dafür sorgen, daß Elektronen von *irgendwoher* in das Leitungsband oder aus dem Valenzband unter Zurücklassung von Löchern *irgendwohin* abgegeben werden können. Dieses «irgendwoher» oder «irgendwohin» sind meist Fremdatome, die man in das reine Halbleitermaterial einbringt und die gegenüber den Atomen des Halbleiters ein Valenzelektron mehr oder weniger haben. Wir geben dafür ein Beispiel (siehe Abb. 7). Silizium ist vierwertig, hat also vier Valenzelektronen. Phosphor ist fünfwertig, hat also fünf Valenzelektronen. Ersetzen wir 1 Silizium- durch 1 Phosphoratom, so wird das 5. Elektron des

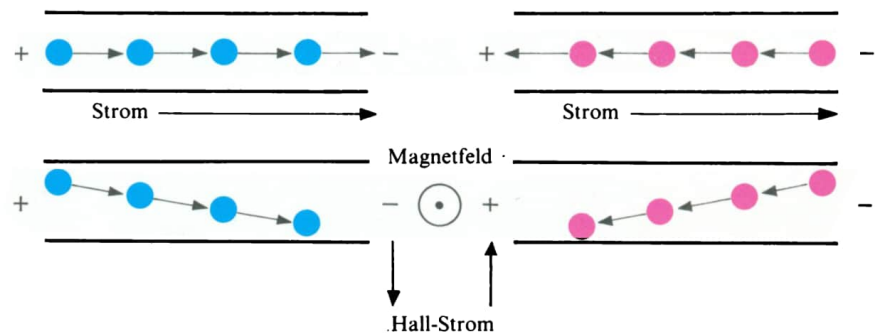


Abb. 6 Hall-Effekt im p- und n-leitenden Halbleiter

a) Leitungsband

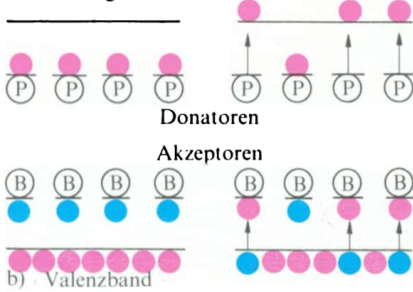


Abb. 7a Entstehung von Elektronen im Leitungsband beim Dotieren des 4wertigen Siliziums mit dem 5wertigen Phosphor. Die meisten Phosphoratome geben ihr 5. Valenzelektron an das Leitungsband des Siliziums ab – sie wirken als Donatoren von Elektronen.

7b Entstehung von Löchern im Valenzband beim Dotieren des 4wertigen Siliziums mit dem 3wertigen Bor. Die meisten Boratome nehmen ein zusätzliches Valenzelektron aus dem Valenzband des Siliziums auf – sie wirken als Akzeptoren von Elektronen. Im Valenzband entstehen Löcher.

Phosphors für die chemische Valenzbindung nicht benötigt und kann an das Leitungsband des Siliziums abgegeben werden. Das Einbringen von Phosphoratomen, das *Dotieren* mit Phosphor, wie man sagt, macht Silizium n-leitend. Die entstehende Konzentration freier Elektronen ist dabei in typischen Fällen etwa gleich der oder wenig kleiner als die Konzentration der Fremdatome. Dotieren wir mit dem dreiwertigen Bor statt mit dem fünfwertigen Phosphor, so entsteht p-leitendes Silizium. Der Grund ist klar: Zur vollen Einbindung in den Siliziumkristall fehlt dem Boratom ein Valenzelektron. Es entnimmt dieses dem Valenzband des Siliziums, wodurch dort ein Loch entsteht. Das Silizium wird p-leitend. Man kann also den Leitungstyp durch Dotieren des Halbleitermate-

rials mit Fremdatomen bewußt einstellen. Allerdings haben wir dabei eine Voraussetzung gemacht, die in Wirklichkeit nicht erfüllt ist. Wir haben nämlich angenommen, daß das Halbleitermaterial vor der Dotierung chemisch vollkommen rein und kristallographisch absolut perfekt war. In Abb. 8 ist ein solcher perfekt gebauter Siliziumkristall im Ausschnitt dargestellt. Er enthält ausschließlich Siliziumatome und ist in den drei Richtungen des Raumes unendlich ausgedehnt. Jedes Siliziumatom ist Mittelpunkt eines geometrisch vollkommenen Tetraeders und wird von vier Nachbaratomen umgeben, die auf den Ecken dieses Tetraeders sitzen. Einen idealen Kristall, wie er in Abb. 8 ausschnittsweise dargestellt ist, gibt es in Wirklichkeit nicht. Jeder reale Kristall **Verunreinigungen und Strukturdefekte** enthält chemische Verunreinigungen und kristallographische Baufehler, sogenannte *Strukturdefekte*, und ist endlich. In den Siliziumkristallstäben, wie sie als Ausgangsmaterial in der Mikroelektronik verwendet werden, findet man als Verunreinigungen u. a. Kohlenstoff, Sauerstoff, Eisen, Kupfer und andere Schwermetallatome. Bei den chemischen Verunreinigungen handelt es sich um punktförmige Störungen. Strukturdefekte von punktförmiger Natur sind u. a. *Vakanzen*. Eine Vakanz liegt dann vor, wenn ein Platz im idealen Kristallgitter, der eigentlich von einem Atom besetzt sein müßte, nicht besetzt, also vakant ist. Neben punktförmigen Defekten gibt es auch linien- und flächenförmig verteilte. Eine Versetzungslinie ist eine Störung, bei der die Besetzung einer kristallographischen Ebene, die

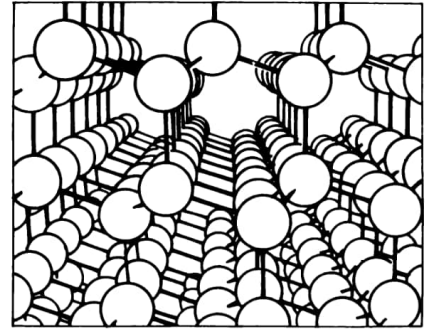


Abb. 8 Struktur eines idealen Siliziumkristalls

eigentlich durchweg mit Atomen besetzt sein müßte, längs einer Markierungslinie plötzlich abbricht. Die Oberfläche eines Halbleiters ist ein flächenförmiger Defekt des im Idealfall unendlich ausgedehnten Kristalls. Jeder reale Kristall enthält eine große Vielfalt derartiger punktförmig, linien- und flächenförmig verteilter Abweichungen von seiner Idealgestalt. Auch wenn ihre Anzahl, gemessen an der ungeheuren Zahl chemisch richtiger, strukturell perfekt angeordneter Atome immer noch verschwindend klein ist (im entgegengesetzten Fall könnte man überhaupt nicht mehr von einem Kristall, auch nicht von einem gestörten, sprechen), haben diese Abweichungen auf die Eigenschaften von Halbleitern doch großen Einfluß. Soweit **Flache und tiefe Niveaus** es die elektrischen Eigenschaften betrifft, beruht dieser Einfluß vor allem auf der Erzeugung von erlaubten Energieniveaus zwischen dem Valenzband und dem Leitungsband. Beim idealen Kristall ist dieser Energiebereich für Elektronen verboten. Beim gestörten Kristall gibt es in ihm erlaubte Energiezustände. Im Unterschied zu den Energiebän-

dern sind die Elektronen oder Löcher in diesen Zuständen aber nicht frei beweglich, sondern an den Ort der Störung gebunden. Man spricht deshalb von lokalisierten Zuständen. Einen bestimmten Typ derartiger Zustände haben wir im Zusammenhang mit den Dotierungsatomen schon kennengelernt. Auch Dotierungsatome sind Störungen und deshalb mit lokalen Energieniveaus innerhalb der Energielücke verbunden. Ihre Besonderheit besteht darin, daß sie relativ flach unterhalb des Leitungsbandes bzw. oberhalb des Valenzbandes liegen (siehe Abb. 7) Man spricht deshalb von *flachen Niveaus*. Aus flachen Niveaus können die Elektronen bzw. Löcher relativ leicht in das benachbarte Leitungs- bzw. Valenzband gelangen und dort den oben beschriebenen Dotierungseffekt hervorrufen. Neben den flachen Niveaus gibt es aber auch sogenannte *tiefe Niveaus*, die sich im mittleren Teil der Energielücke befinden und von den beiden Bändern relativ weit entfernt sind. Fast alle der oben genannten Störungen führen zu tiefen Niveaus. Wegen ihres zu großen Abstandes von den Energiebändern tragen sie zur Dotierung nicht bei. Sie sind, wie man sagt, elektrisch nicht aktiv. In vielen Fällen sind sie einer Dotierung sogar hinderlich. Nehmen wir z. B. an, daß ein Siliziumkristall pro Kubikzentimeter 10^{16} Störungen (chemische Verunreinigungen oder Strukturdefekte) enthält. Mit jeder Störung sei ein tiefes Niveau verbunden, in dem Platz für ein zusätzliches Elektron sei. Versuchen wir, den Siliziumkristall durch die Zugabe von 10^{16} Phosphoratomen pro cm^3 n-leitend zu machen, werden wir eine

Enttäuschung erleben. Statt an das Leitungsband werden die Phosphoratome ihr fünftes Elektron vorwiegend an die tiefen Niveaus abgeben. Dort sitzen die Elektronen fest und sind für die elektrische Leitung unwirksam. Chemische Verunreinigungen und Strukturdefekte beeinflussen die Dotierbarkeit eines Siliziumkristalls also u. U. bereits dann wesentlich, wenn sie in einer Konzentration von 10^{16} pro cm^3 auftreten. Bei 10^{22} Atomen pro cm^3 insgesamt sind das nur 0,0001 %. Vielleicht waren aber im Ausgangsmaterial nicht nur Störungen vorhanden, die zu tiefen Niveaus Anlaß geben, sondern auch Phosphor- oder Boratome in einer Konzentration bei oder über 10^{16} pro cm^3 . Auch in diesem Falle ändert sich die Wirkung der nachträglich und absichtlich zwecks Dotierung eingebrachten Phosphoratome wesentlich. Was wir hier am speziellen Beispiel, und bezogen auf die Dotierbarkeit, dargestellt haben, gilt ganz allgemein und nicht nur aus Gründen der Dotierbarkeit: An Halbleitermaterialien müssen extrem hohe Anforderungen bezüglich chemischer Reinheit und kristallographischer Perfektion gestellt werden. Die gezielte Zugabe kleinster Verunreinigungsmengen muß beherrscht werden. In der Chemie ist ein Material mit 0,001 % Verunreinigungen schon hochrein. In der Halbleiterphysik und -technik werden 100- und 1000-fach reinere Materialien benötigt. «Halbleiterrein» ist zum Inbegriff für extreme Reinheit geworden. Zu den reinsten Halbleitermaterialien, die bisher hergestellt wurden, gehört das Germanium mit einer Konzentration von 10^9 elektrisch aktiven

Fremdatomen pro cm^3 . Siliziumkristalle lassen sich ebenfalls sehr rein und perfekt züchten. Außerdem ist beim Silizium die elektrische Wirkung solcher unvermeidlicher Defekte, wie sie eine Oberfläche oder genauer gesagt, die Grenzfläche zum Siliziumdioxid darstellt, relativ gering. Das sind einige der Gründe dafür, warum in der Mikroelektronik ausgerechnet Silizium als Halbleitermaterial verwendet wird. Auf andere Gründe werden wir bei der Behandlung der Herstellungstechnologie der Mikroelektronik in Kapitel 5 noch eingehen. Die in der Natur vorkommenden Kristalle erreichen auch nicht annähernd den Grad chemischer Reinheit und kristallographischer Perfektion, wie sie für Halbleiter erforderlich sind. Kaum einer der natürlich vorkommenden Halbleiterkristalle ist dotierbar und nur wenige sind überhaupt halbleitend, wie Bleiglanz oder bestimmte natürliche Diamanten. Halbleiterkristalle sind also höchst künstliche Materialien. Bei ihrer Herstellung muß ein hoher Aufwand getrieben werden. Davon wird in Kapitel 5 im Zusammenhang mit der Herstellungstechnologie der Mikroelektronik noch die Rede sein.

Erzeugung

von Überschußladungsträgern

Als Grundlage für die Informationsverarbeitung mit Hilfe der Elektronik hatten wir weiter vorn herausgestellt, daß die Bewegung der Elektronen bzw. die Anzahl der beweglichen Elektronen durch äußere Einwirkungen beeinflußt werden kann. Wie geschieht diese Beeinflussung nun konkret bei einem Halbleiter? Sie beruht auf der Möglichkeit, über die bereits im Normalzustand im

Halbleiter vorhandene Anzahl von Elektronen und Löchern hinaus eine merkliche Zahl *zusätzlicher* Ladungsträger zu erzeugen. Wir betrachten diese Generation von Überschubladungsträgern zunächst an einem einfachen Beispiel. Wird ein Halbleiter mit Licht passender Wellenlänge bestrahlt, so werden Elektronen aus dem Valenzband unter Absorption eines Lichtquants und Zurücklassung eines Loches in das Leitungsband angehoben. Es entstehen zusätzliche Elektronen und Löcher (Abb. 9a). Die elektrische Leitfähigkeit des Halbleiters erhöht sich — allerdings nur für einen kurzen Moment, denn die Lebensdauer der überschüssigen Elektronen und Löcher ist begrenzt. Dafür sorgen Prozesse im Halbleiter, die solange nicht zur Ruhe kommen, bis die Zahl der Elektronen und Löcher auf einen Wert gebracht ist, der dem Normalzustand, dem sogenannten thermodynamischen Gleichgewicht, entspricht. In Abb. 9 sind einige dieser Prozesse dargestellt. Elektronen können z. B. direkt in das Loch zurückfallen (Abb. 9b). Man spricht

Rekombination

dann von der Rekombination eines Elektron-Loch-Paares. Ein Elektron oder ein Loch kann aber auch einzeln von einem tiefen Niveau eingefangen werden, an dem es danach haften bleibt (Abb. 9c, d). Fängt dasselbe tiefe Niveau nacheinander ein Elektron und ein Loch ein, ist das Ergebnis das gleiche wie bei der Rekombination: ein Elektron-Loch-Paar ist verschwunden (Abb. 9e). Oft ist es gerade dieser Prozeß, die Rekombination über tiefe Niveaus, der die Lebensdauer von Überschubladungsträgern in Halbleitern begrenzt.

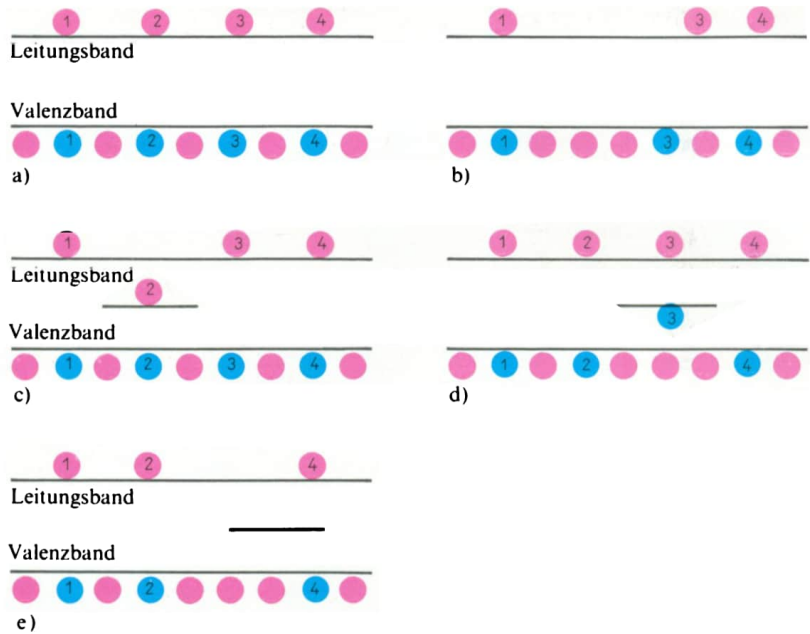


Abb. 9 Prozesse, die zur Vernichtung von Überschub-Ladungsträgern führen:

a) Überschub-Elektronen im Leitungsband und Überschub-Löcher im Valenzband

b) Rekombination: Elektron 2 ist in das Loch 2 hineingefallen, beide sind verschwunden — sie haben sich gegenseitig vernichtet. Die dabei freiwerdende Energie wird an den Kristall abgegeben oder als Licht emittiert.

c) Einfang eines Elektrons durch ein tiefes Niveau: Elektron 2 ist als frei bewegliches Elektron des Leitungsbandes verschwunden. An der Zahl der Löcher hat sich nichts geändert.

d) Einfang eines Loches durch ein tiefes Niveau: Loch 3 ist als frei bewegliches Loch des Valenzbandes verschwunden. An der Zahl der Elektronen hat sich nichts geändert.

e) Rekombination über ein tiefes Niveau: Elektron 3 und Loch 3 sind beide von demselben tiefen Niveau eingefangen worden.

Wünscht man eine lange Lebensdauer, muß man die Zahl der tiefen Niveaus gering halten, d. h., man muß mit reinen und kristallographisch perfekten Materialien arbeiten. Ebenso wie weiter oben die Forderung nach guter Dotierbarkeit läuft also auch die Forderung nach langer Lebensdauer von Überschubladungsträgern auf eine hohe chemische Reinheit und kristallografische Perfektion hinaus. Über kurz oder lang verschwinden aber auch im besten Halbleitermaterial die einmal erzeugten Überschubladungsträger.

Wenn nicht durch anhaltende Bestrahlung mit Licht ständig wieder neue Elektronen-Loch-Paare nachgeliefert werden, stellt sich keine dauernde Änderung der elektrischen Leitfähigkeit ein. Sie klingt auf die Dunkelleitfähigkeit ab.

pn-Übergang. Halbleiterdiode

Die soeben besprochene Generation von Überschubladungsträgern durch Licht wird allerdings in der Mikroelektronik nicht angewandt. Wir haben sie kurz behandelt, weil sie der einfachste Mechanismus zur Erzeugung von Überschubladungsträ-

gern ist. Für die Mikroelektronik ist ein anderer Mechanismus von Bedeutung, auf den wir jetzt zu sprechen kommen. Zusätzliche Elektronen in einem p-Halbleiter, in dem von Natur aus viele Löcher, aber wenige Elektronen vorhanden sind, kann man auch dadurch schaffen, daß man ihn in Kontakt mit einem n-Halbleiter bringt, der von Natur aus viele Elektronen enthält. Man spricht von einem *pn-Übergang*. Die Elektronen *diffundieren* an einem solchen pn-Übergang in den p-Halbleiter, d. h., sie dringen vermöge ihrer thermischen Bewegung in den p-Halbleiter vor, ähnlich wie die Moleküle eines Gases in ein evakuiertes Gefäß eindringen, wenn ihnen der Weg dazu freigegeben wird. Der p-Halbleiter ist für die Elektronen des n-Halbleiters ein Vakuum. Im Unterschied zu den Gasmolekülen haben die Elektronen im p-Halbleiter aber nur eine geringe Lebensdauer. Sie rekombinieren mit den Löchern des p-Halbleiters. Ihre Anzahl wird um so kleiner, je länger sie sich im p-Halbleiter aufhalten, d. h. insbesondere auch, je weiter sie sich vom n-Halbleiter entfernen. Durch die Abwanderung von Elektronen entsteht im n-Halbleiter in der Nähe der Grenzfläche zum p-Halbleiter ein von freien Ladungsträgern verarmtes, positives Raumladungsgebiet. Man nennt es auch *Verarmungsgebiet*. Um Mißverständnissen zu begegnen, bemerken wir, daß die positive Raumladung im Verarmungsgebiet nicht etwa von den positiven Löchern herrührt (Löcher gibt es auf der n-Seite des pn-Überganges praktisch überhaupt nicht), sondern von den Dotierungsatomen im Halbleiter, die positiv geladen

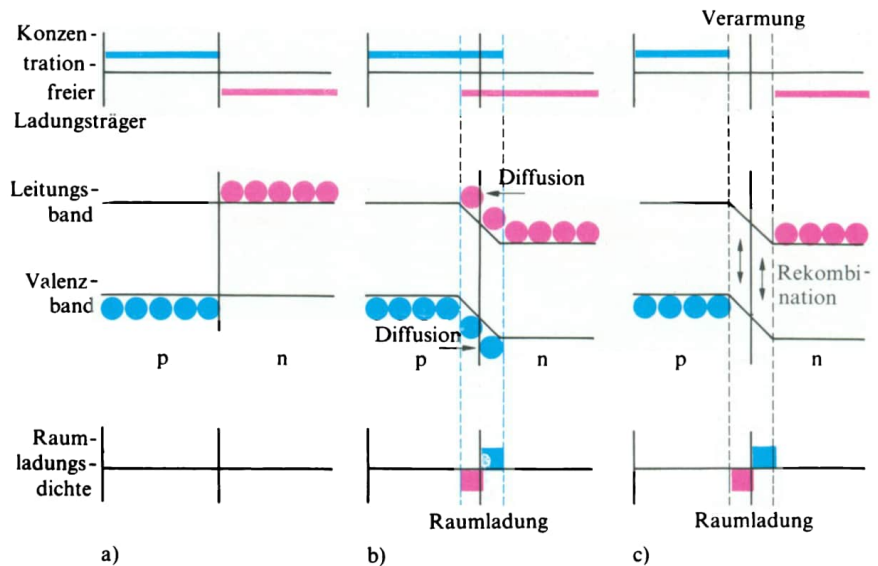


Abb. 10 Einstellung des Gleichgewichts an einem pn-Übergang:

a) Unmittelbar nach Herstellung des Kontaktes zwischen p- und n-Gebiet.

b) Die Elektronen diffundieren in das p-Gebiet und die Löcher in das n-Gebiet. Dabei baut sich eine Raumladung auf, die im p-Gebiet negativ und im n-Gebiet positiv ist.

c) Elektronen und Löcher rekombinieren im Raumladungsgebiet, es bildet sich ein Verarmungsgebiet für freie Ladungsträger. Damit ist der Gleichgewichtszustand am pn-Übergang erreicht.

sind, wenn von ihnen Elektronen entfernt werden, so wie es im Verarmungsgebiet geschehen ist. Die sich aufbauende Raumladung im Verarmungsgebiet erschwert es weiteren Elektronen zunehmend, den n-Halbleiter zu verlassen, bis schließlich nur noch soviel Elektronen nachgeliefert werden, wie zur Aufrechterhaltung der bereits vorhandenen Raumladung gegen die abbauende Wirkung der Rekombination erforderlich sind. Was wir hier über das Eindringen der Elektronen des n-Halbleiters in den p-Halbleiter festgestellt haben, trifft in analoger Weise auf die Löcher des p-Halbleiters zu. Auch sie diffundieren in das angrenzende Vakuum, in diesem Falle das n-Gebiet, rekombinieren dort teilweise und werden durch eine

sich im p-Gebiet aufbauende negative Raumladung zunehmend am Übertritt in den n-Halbleiter gehindert. An einem pn-Übergang findet also ein relativ kompliziertes Wechselspiel zwischen Diffusion, Rekombination und Aufbau einer Raumladung statt. Der Gleichgewichtszustand, der sich einstellt, wenn sich diese drei Faktoren ausbalanciert haben, ist in Abb. 10 schematisch dargestellt.

Injektion

Wir wollen jetzt diesen Gleichgewichtszustand dadurch stören, daß wir eine Spannung an den pn-Übergang anlegen. Was wird passieren? Klar ist, daß ein Strom fließen wird. Aber auf welche Weise geschieht das? Auf eine recht merkwürdige Art und Weise! Im p-Gebiet, weitab vom

Raumladungsgebiet, ist der Strom hauptsächlich ein Löcherstrom — bei positiver Polung des p-Gebietes fließen die Löcher nach rechts in das Raumladungsgebiet hinein und diffundieren von dort weiter ins n-Gebiet (siehe Abb. 11). Man spricht von *Ladungsträger-Injektion*. Im n-Gebiet treffen die Löcher auf die dort in großer Zahl vorhandenen Elektronen und rekombinieren mit diesen. Es müssen Elektronen aus der Tiefe des n-Gebietes nachgeliefert werden, um das durch die Rekombination entstandene Elektronendefizit auszugleichen. So viele Löcher und Elektronen fließen pro Sekunde in das Rekombinationsgebiet hinein, wie dort pro Sekunde rekombinieren. Was wir über die aus dem p-Gebiet ins n-Gebiet gelangenden Löcher ausgesagt haben, gilt ganz analog auch für Elektronen, die aus dem n-Gebiet ins p-Gebiet vordringen. Sie rekombinieren mit den Löchern im p-Gebiet. Beide Stromanteile tragen zum Gesamtstrom bei. Der Strom ist ein Rekombinationsstrom. Die Stromstärke ist durch die Rekombinationsgeschwindigkeit von Elektronen und Löchern bestimmt. Die Rekombination erfolgt relativ schnell, da dazu keine Energie erforderlich ist, sondern sogar noch Energie abgegeben werden kann. Deshalb ist der Rekombinationsstrom relativ stark. Der pn-Übergang ist in Durchlaßrichtung gepolt. Anders ist es bei negativer Polung des p-Gebietes. In diesem Fall fließen die Löcher im p-Gebiet nach links, die Elektronen im n-Gebiet nach rechts. Es können pro Sekunde so viele Löcher nach links und Elektronen nach rechts abfließen, wie am pn-Übergang pro Sekunde generiert werden. Der Strom

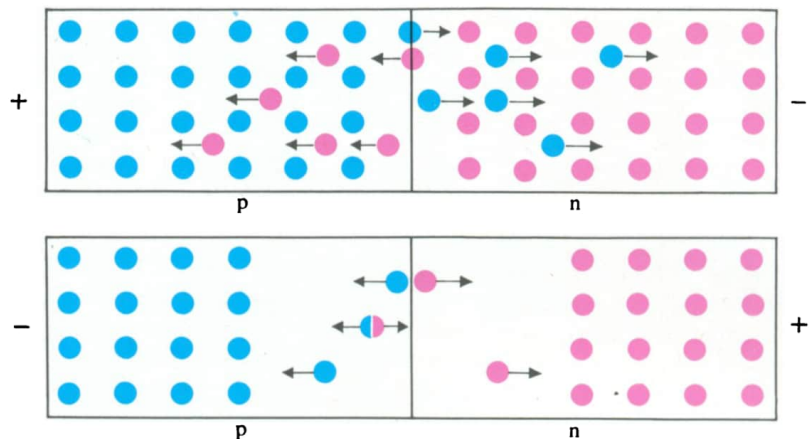


Abb. 11 Stromtransport durch einen pn-Übergang

a) In Durchlaßrichtung. Die aus dem p-Gebiet in das n-Gebiet vordringenden Löcher und die aus dem n-Gebiet in das p-Gebiet vordringenden Elektronen rekombinieren dort. Der Strom ist ein Rekombinationsstrom von Majoritätsladungsträgern.

b) In Sperrrichtung. Die im n-Gebiet nach rechts abfließenden Elektronen und die im p-Gebiet nach links abfließenden Löcher müssen durch Generation aus dem Raumladungsgebiet nachgeliefert werden. Der Strom ist ein Generationsstrom von Majoritätsladungsträgern.

ist also ein Generationsstrom. Die Generation erfolgt relativ langsam, weil dazu Energie aufgebracht werden muß. Daher ist der Generationsstrom relativ schwach. Der pn-Übergang ist in Sperrrichtung gepolt. Er wirkt also wie ein Gleichrichter. Man bezeichnet ihn deshalb auch als *Halbleiter-Gleichrichter* oder als Halbleiterdiode. Wie der Ladungsstrom in einem Stromkreis mit einer solchen Halbleiterdiode von den Elektronen auf die Löcher und von den Löchern auf die Elektronen übergeht, ist in Abb. 12 schematisch dargestellt. In den metallischen Leitungen, die die Diode mit der Spannungsquelle verbinden, wird der Strom natürlich von Elektronen getragen. An den beiden Kontaktstellen zwischen dem metallischen Leiter und den beiden Halbleitergebieten entsteht ebenfalls eine Raumladung.

und des Halbleitermaterials kann es sich dabei um eine Verarmung von freien Ladungsträgern, wie im Fall des pn-Übergangs, oder um eine Anreicherung handeln. Im Fall der Anreicherung ist der Kontakt gut leitend, im Fall der Verarmung schlecht. In der Mikroelektronik ist man an gut leitenden, sogenannten «Ohmschen» Metall-Halbleiter-Kontakten interessiert. Die Kombinationen zwischen Aluminium und

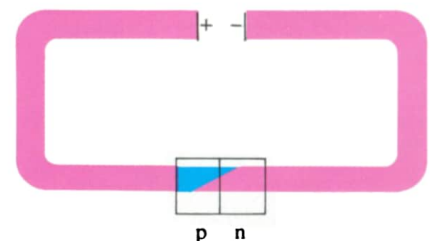


Abb. 12 Elektronen- und Löcherfluß in einem Stromkreis mit pn-Übergang. Es sind nur die Majoritätsladungsträgerströme dargestellt.

stark dotiertem p- oder n-Silizium ergeben gutleitende Kontakte. Auch deshalb wird meist Aluminium als Kontaktmaterial verwendet.

Bipolarer Transistor

Die Wirkungsweise eines *bipolaren Transistors* läßt sich auf der gleichen Basis verstehen wie der pn-Gleichrichter. Beim npn-Transistor befindet sich ein n-Gebiet im Kontakt mit einem p-Gebiet, das seinerseits mit einem zweiten n-Gebiet im Kontakt steht (siehe Abb. 13). Wesentlich ist dabei, wie wir gleich noch sehen werden, daß das mittlere p-Gebiet sehr schmal und relativ niedrig dotiert ist. Beim pnp-Transistor sind p und n gegenüber dem npn-Transistor vertauscht. Beide Arten von Transistoren werden in der Praxis verwendet und gleichen sich in ihrer Funktion, nur daß beim pnp-Transistor der Strom hauptsächlich durch Löcher getragen wird, beim npn-Transistor hauptsächlich durch Elektronen. Da die Elektronen in Silizium beweglicher sind als die Löcher, ist der npn-Transistor dem pnp-Transistor bei den meisten Anwendungen überlegen. Wir erläutern die Wirkungsweise für den Fall des npn-Transistors aus Abb. 13. Er besitzt zwei pn-Übergänge, genauer links einen np- und rechts einen pn-Übergang. Wir legen an das rechte n-Gebiet eine hohe positive Spannung U_K und an das linke n-Gebiet eine negative Spannung U_E an. Das p-Gebiet wird an eine Spannung U_B gelegt, die größer ist als U_E , aber kleiner als U_K . Damit ist der linke np-Übergang in Durchlaßrichtung, der rechte pn-Übergang in Sperrrichtung gepolt. Die Elektronen treten also aus dem linken n-Gebiet aus — es wird daher

als *Emitter* bezeichnet — und diffundieren in das p-Gebiet, wo sie normalerweise rekombinieren würden, wenn das p-Gebiet — die sogenannte *Basis* — nicht so schmal gemacht worden wäre, daß die Elektronen schon hindurch diffundiert sind, noch bevor sie Zeit hatten zu rekombinieren. Durch diesen Kunstgriff wird erreicht, daß die Elektronen in das rechte n-Gebiet gelangen und dort von der positiven Elektrode abgesaugt werden können. Das rechte n-Gebiet bezeichnet man daher als *Kollektor*. Es sammelt die aus dem linken n-Gebiet, dem Emitter, emittierten Elektronen auf. Die unterschiedliche Rolle der beiden n-Gebiete, die ohne Spannung völlig gleichberechtigt sind, ergibt sich

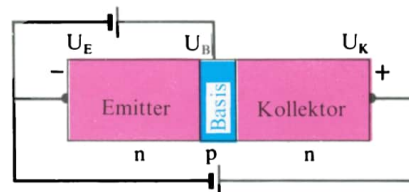


Abb. 13 npn-Transistor. Gezeigt ist die Emitterschaltung.

natürlich daraus, daß an das rechte n-Gebiet eine positive und an das linke eine negative Spannung angelegt wurde. Nun verbinden wir die Basis über eine Spannungsquelle leitend mit dem Emitter, so wie das in Abb. 13 bereits geschehen ist. Wir haben jetzt zwei Stromkreise, einen Emitter-Basis-Kreis und einen Emitter-Kollektor-Kreis. Zum Strom im Emitter-Basis-Kreis tragen nur diejenigen Elektronen bei, die — wie wir von der Behandlung des Stromflusses durch den in Durchlaßrichtung gepolten pn-Übergang wissen —

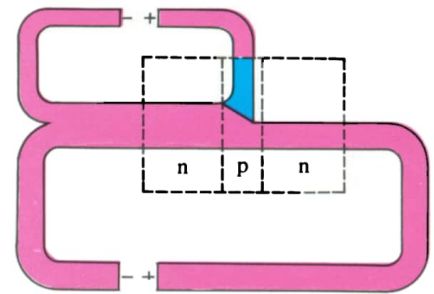


Abb. 14 Elektronen- und Löcherfluß in einem Stromkreis mit npn-Transistor in Emitterschaltung. Es sind wieder nur die Majoritätsladungsträgerströme dargestellt.

im p-Gebiet rekombinieren. Das sind aber wegen der Schmalheit des p-Gebietes sehr wenig. Der über die Basis abfließende Elektronenstrom ist also sehr gering. Der größte Teil des aus dem Emitter kommenden Stromes fließt über den Kollektor ab (siehe Abb. 14). Wie wir wissen, tragen zum Stromfluß durch einen pn-Übergang aber nicht nur Elektronen, sondern auch Löcher bei. Das heißt in unserem Falle, daß aus der Basis in den Emitter ein Löcherstrom fließt. Er würde den relativ kleinen Elektronenstrom zwischen Emitter und Basis wesentlich übertreffen und insgesamt zu einem relativ großen Strom im Emitter-Basis-Kreis Anlaß geben, wenn man ihn nicht durch einen weiteren Kunstgriff sehr klein machen könnte. Dieser zweite Kunstgriff beim bipolaren Transistor ist die besonders niedrige Dotierung der Basis. Dadurch ist die Löcherkonzentration in der Basis gering, und es kann nur ein kleiner Löcherstrom aus ihr abfließen. Die oben für den Elektronenstrom getroffene Feststellung bleibt also auch für den Gesamtstrom aus Elektronen und Löchern richtig: Der größte Teil des Stromes durch den Emitter

fließt im Emitter-Kollektor-Kreis und nur ein geringer Teil im Emitter-Basis-Kreis.

Welchen Nutzen kann man nun daraus ziehen? Ändern wir die Spannung zwischen Emitter und Basis bei konstanter Spannung zwischen Basis und Kollektor, so ändert sich der aus dem Emitter herausfließende Strom. Der größte Teil dieser Stromänderung entfällt dabei auf den Emitter-Kollektor-Kreis, der geringere auf den Emitter-Basis-Kreis. Mit Hilfe einer kleinen Stromänderung im Emitter-Basis-Kreis läßt sich also eine große Stromänderung im Emitter-Kollektor-Kreis bewirken. Die Stromänderung wird verstärkt. Da auch der Widerstand des Emitter-Basis-Kreises wesentlich kleiner ist (der pn-Übergang ist in Durchlaßrichtung gepolt) als der Widerstand des Emitter-Kollektor-Kreises (ein pn-Übergang ist in Sperrichtung gepolt), fällt auch die im Emitter-Basis-Kreis umgesetzte elektrische Leistung wesentlich kleiner aus als die im Emitter-Kollektor-Kreis. Mit Hilfe einer kleinen elektrischen Leistung im Emitter-Basis-Kreis läßt sich also eine große Leistung im Emitter-Kollektor-Kreis steuern. Die gleiche Funktion erfüllt in der Vakuumelektronik eine Triodenröhre. Der Emitter des Transistors entspricht dabei der Kathode der Elektronenröhre, der Kollektor der Anode und die Basis dem Gitter. Die in Abb. 13 dargestellte Schaltung des Transistors, bei der der Emitter mit den beiden übrigen Gebieten, der Basis und dem Kollektor verbunden ist, während die Basis und der Kollektor jeweils nur mit einem Gebiet, eben dem Emitter, verbunden sind, nennt man Emitter-

schaltung. Sie ist eine von drei möglichen Schaltungsarten des Transistors. Bei der Kollektorschaltung ist der Kollektor das gemeinsame Gebiet für die beiden Stromkreise, bei der Basisschaltung ist es die Basis. In der Mikroelektronik werden auch Transistoren verwendet, die zwei oder mehrere voneinander isolierte Emittergebiete besitzen. Die Schaltungsmöglichkeiten werden dadurch reichhaltiger. Auch Transistoren mit mehreren Kollektoren sind möglich.

Mit Hilfe des Transistors kann man elektrische Leistungen verstärken oder schalten. In der Mikroelektronik wird der Transistor hauptsächlich in seiner Eigenschaft als ein durch elektrische Impulse betätigbarer Schalter eingesetzt. Eine wichtige Kenngröße für derartige Schalter ist die für einen Schaltvorgang «ein-aus» oder «aus-ein» erforderliche Zeit, kurz Verzögerungszeit genannt. Ein Schalter ist um so besser, je kürzer die Verzögerungszeit ist. Um eine Vorstellung von der Größe der Verzögerungszeit beim bipolaren Transistor zu erhalten, stellen wir die folgende Überlegung an: Emitter, Basis und Kollektor eines npn-Transistors seien an eine bestimmte elektrische Spannung U_E , U_B und U_K gelegt. Im Emitter-Kollektor-Kreis fließt ein bestimmter Strom. Wird die Spannung der Basis auf einen Wert $U_B + \Delta U_B$ erhöht, so nimmt die Durchlaßfähigkeit sowohl des linken np-Übergangs als auch die des rechten pn-Übergangs zu und der Strom im Emitter-Kollektor-Kreis wird größer. Wir haben dabei aber stillschweigend angenommen, daß während des Hindurchdiffundierens der Elektronen durch

die Basis die Spannung der Basis unverändert gleich $U_B + \Delta U_B$ geblieben ist. Die Verhältnisse ändern sich, wenn die Basisspannung während dieser Zeit nicht konstant bleibt. Ist sie z. B. schon wieder auf den Wert U_B zurückgegangen, bevor die Elektronen Zeit hatten, durch die Basis zu diffundieren, so ist der durch ΔU_B bewirkte Stromzuwachs im Emitter-Kollektor-Kreis nicht so groß wie bei einer konstanten Spannung $U_B + \Delta U_B$. Wir schließen daraus, daß sich die Schaltwirkung des Transistors verschlechtert, wenn die Schaltbefehle in Form höherer oder niedrigerer Spannungen an der Basis in zeitlichen Abständen aufeinanderfolgen, die kleiner sind als die Zeit, die die Elektronen zum Diffundieren durch die Basis benötigen. Die Verzögerungszeit T des Transistors ist also die Zeit zum Diffundieren durch die Basis. Es ist klar, daß diese Zeit um so kleiner sein wird, je schmaler die Basis ist. Daraus ergibt sich das Bestreben der Mikroelektronik, vor allem die Basis möglichst schmal zu machen. Wir haben damit auch die in Kapitel 2 aufgestellte Behauptung bewiesen, daß die Bauelemente schneller werden, wenn man sie nur kleiner macht. Die Größe der Diffusionszeit T durch die Basis läßt sich leicht abschätzen, wenn wir den Diffusionskoeffizienten der Elektronen D und die Breite der Basis b kennen. Allgemein gilt $T \sim b^2/D$. Rechnet man für Silizium mit einer Basisbreite von $0,7 \mu\text{m}$ und einem Diffusionskoeffizienten der Elektronen von $50 \text{ cm}^2 \text{ s}^{-1}$, so folgt für die Verzögerungszeit des npn-Transistors $T \approx 10^{-10} \text{ s}$. Der Diffusionskoeffizient der im Vergleich zu den Elektronen nicht so

beweglichen Löcher des Siliziums beträgt nur rund 1/3 des Wertes für die Elektronen. Die Verzögerungszeit des pnp-Transistors ist deshalb dreimal so groß wie beim npn-Transistor. Noch beweglicher als im Silizium sind die Elektronen in Galliumarsenid (GaAs). Aus n-GaAs sollten sich daher besonders schnelle npn-Transistoren herstellen lassen.

Wir sind jetzt auch in der Lage, die für einen Schaltvorgang in einem npn-Transistor erforderliche Energie auszurechnen. Dazu müssen wir die für einen Schaltakt erforderliche Zeit T mit der dabei verbrauchten elektrischen Leistung P multiplizieren. Bei einem bipolaren Transistor wird P im Emitter-Basis-Kreis umgesetzt. Allgemein ergibt sich die Leistung als Produkt aus Spannung und Stromstärke. In unserem konkreten Fall hängt die Stromstärke von der Größe des Transistors ab. Je kleiner der Transistor ist, um so geringere Ströme treten in ihm auf. In integrierten Schaltkreisen sind die Transistoren auf einer Fläche von $1000 \mu\text{m}^2$ und darunter untergebracht. Im Emitter-Basis-Kreis fließen Ströme von der Größe einiger Mikroampere oder darunter, die Spannung beträgt etwa 1 Volt. Die zum Schalten erforderliche Leistung ist also gleich 1 Volt mal einige Mikroampere, d. h. einige Mikrowatt. Mit $P = 10^{-5} \text{ W}$ und $T = 10^{-10} \text{ s}$ ergibt sich für die Schaltenergie $10^{-5} \times 10^{-10} \text{ Ws} = 10^{-15} \text{ Joule}$. Das ist eine außerordentlich kleine Energiemenge, die allerdings nichts mit dem wirklichen Energieverbrauch eines Transistors zu tun hat. Der Energieverbrauch im Emitter-Kollektor-Kreis ist wesentlich größer.

Feldeffekttransistor. MOS-FET
Einfacher als die Wirkungsweise des bipolaren Transistors, den wir soeben besprochen haben, ist die des *Feldeffekttransistors*. Beide Typen von Transistoren werden heute in mikroelektronischen Schaltkreisen verwendet. Der Feldeffekttransistor kommt dabei sogar noch häufiger zum Einsatz als der bipolare Transistor, insbesondere in hochintegrierten Schaltkreisen wie den Halbleiterspeichern. Die Erfindung des Feldeffekttransistors ist auch wissenschaftsgeschichtlich interessant. Er wurde mindestens dreimal erfunden. Beschrieben, aber nicht realisiert, wurde er 1928 von Lillienfeld und 1930 von Heil. 1947 baute Shockley einen Feldeffekttransistor, der aber völlig unzureichend funktionierte. Bei der Suche nach den Ursachen dieses Mißerfolges erfanden Bardeen und Brattain, gewissermaßen als Nebenprodukt, 1948 den bipolaren Transistor. Der Feldeffekttransistor konnte erst in den sechziger Jahren verwirklicht werden, nachdem es durch Verbesserung der Kristallzüchtungsverfahren gelungen war, wesentlich perfektere Siliziumkristalle herzustellen und nachdem die Entwicklung der Planar-Technik geeignete Bauelementstrukturen möglich gemacht hatte. Das Prinzip des Feldeffekttransistors ist in Abb. 15 dargestellt. Eine p- oder n-leitende Halbleiterprobe (im folgenden nehmen wir eine n-leitende an) wird auf einer Seite mit einem Isolator versehen und zwischen zwei Metallelektroden gebracht. Wird an die obere Elektrode eine positive Spannung angelegt, so werden in den Halbleiter Elektronen hineingezogen, man sagt, es werden Ladungen *in-*

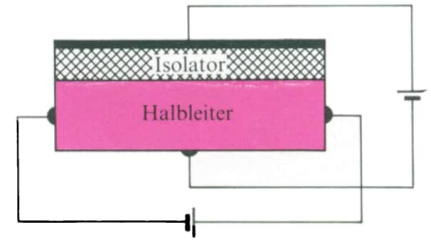


Abb. 15 Prinzip des Feldeffekttransistors

fluenziert. Ist die Zahl der influenzierten Elektronen vergleichbar mit der Zahl der freien Elektronen, die bereits ohne Spannung im Halbleiter vorhanden sind, so wird dabei dessen Widerstand wesentlich verändert. Der Strom im oberen Kreis in Abb. 15 ist praktisch Null und damit auch die in ihm verbrauchte elektrische Leistung, während im unteren Kreis ein stärkerer Strom fließen und eine größere elektrische Leistung umgesetzt werden kann. Man hat also wieder die Möglichkeit, mit Hilfe einer geringeren elektrischen Leistung eine wesentlich größere zu steuern, also zu verstärken bzw. auf elektrischem Wege zu schalten. Dieses Prinzip ist so einleuchtend und einfach wie seine Verwirklichung schwierig ist. Auf zwei Probleme wollen wir kurz eingehen. Eine positive Spannung anzulegen heißt, auf die Oberfläche des Isolators positive elektrische Ladung zu bringen — je mehr Ladung, um so größer die Spannung. Wenn wir pro Atom der ersten Atomschicht des Isolators ein Elektron wegnehmen (d. h. eine positive Elementarladung dazugeben), werden wir eine Feldstärke erzeugen, die etwa von der Größe der elektrischen Felder im Atom ist, also etwa 10^8 V/cm . Dazu sind bei einem Abstand zwischen den Atomen in der Oberflächenschicht

von $3 \cdot 10^{-8} \text{ cm}$ etwa 10^{15} cm^{-2} Elektronen zu entfernen. Genauso viele Elektronen werden im Halbleiter influenziert. Denken wir uns ein Halbleiterplättchen von 1 mm Dicke, so befinden sich unter 1 cm^2 Oberfläche bereits ohne angelegte Spannung etwa 10^{15} Elektronen, falls die Ladungsträgerkonzentration des Halbleiters 10^{16} cm^{-3} beträgt, was ein typischer Wert ist. Um zu erreichen, daß wenigstens genausoviel Elektronen influenziert werden, wie bereits im Halbleiter freie Elektronen vorhanden sind, müßte man also die riesige Spannung von 10 Millionen Volt anlegen. Aber selbst, wenn das gelänge, wäre das Ergebnis enttäuschend. Die meisten der influenzierten Elektronen würden nämlich von den bereits früher erwähnten Störniveaus eingefangen werden, und nur wenige würden frei beweglich bleiben und zur Widerstandsänderung beitragen. Auf diesem Wege ist der Feldeffekttransistor also kaum zu realisieren. Der wirkliche Feldeffekttransistor ist deshalb etwas anders aufgebaut. In Abb. 16 ist er schematisch dargestellt. Charakteristisch an der Anordnung in Abb. 16 sind die beiden spiegelbildlich zueinander gelegenen pn-Übergänge. Wird zwischen dem linken und dem rechten n-Gebiet eine Spannung angelegt, so ist unabhängig von deren

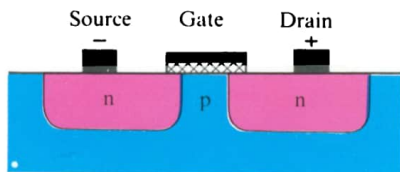


Abb. 16 Schematischer Aufbau eines Feldeffekttransistors, wie er in der Mikroelektronik verwendet wird.

Vorzeichen stets einer der beiden pn-Übergänge gesperrt. Es fließt also in jedem Falle ein sehr geringer Strom. Der Transistor befindet sich im gesperrten Zustand. Leitend kann er mit Hilfe einer dritten Elektrode, dem sogenannten Gate, gemacht werden. Die Gate-Elektrode ist, durch eine dünne Isolatorschicht getrennt, über dem p-Gebiet zwischen den beiden n-Gebieten angeordnet und dient zur Influenzierung von Elektronen, wie sie im Zusammenhang mit der Feldeffektanordnung in Abb. 15 besprochen wurde. Nur werden diese Elektronen jetzt in einen *p-Halbleiter* influenziert, der außerdem ein von Löchern verarmtes Raumladungsgebiet enthält. Man wird eine wesentlich größere Wirksamkeit der influenzierten Elektronen erwarten. Allerdings bleiben die ersten influenzierten Elektronen auch in diesem Fall für den Ladungstransport unwirksam; sie rekombinieren mit den Löchern im p-Gebiet, die zwischen den beiden pn-Gebieten noch vorhanden sind. Sind alle Löcher kompensiert, führt die weitere Influenzierung zu einem Überschuß an Elektronen im p-Gebiet an der Grenzfläche zum Isolator. Es bildet sich ein dünner n-leitender Kanal zwischen den beiden n-Gebieten links und rechts aus. Die beiden n-Gebiete sind jetzt durch einen n-leitenden Kanal miteinander verbunden. Die Elektronen können durch ihn ungehindert vom linken n-Gebiet in das rechte fließen, falls das rechte gegenüber dem linken auf ein positives Potential gelegt wird. Man nennt das linke n-Gebiet dann *Source* bzw. Quelle und das rechte *Drain* bzw. Senke. Das Vorhandensein eines n-leitenden Gebietes in

einem p-Halbleiter muß auf den ersten Blick verwundern, ist aber leicht zu verstehen: p-leitend ist der Halbleiter nur im thermodynamischen Gleichgewicht, ohne ein äußeres elektrisches Feld. Wird, wie das durch die Gate-Elektrode geschieht, im p-Halbleiter ein genügend starkes Feld erzeugt, so werden die Energiebänder abgesenkt und ihre Besetzung ändert sich. Die Löcher im Valenzband werden vollständig mit Elektronen aufgefüllt und auch das Leitungsband wird zum Teil mit Elektronen besetzt. Man nennt diesen Vorgang *Inversion*. Der n-Kanal im p-Halbleiter ist eine Inversionsschicht. Die Spannung am Gate, die zur Ausbildung der Inversionsschicht ausreicht, nennt man *Schwellspannung*. Wird die Schwellspannung überschritten, schaltet der Transistor aus dem gesperrten in den leitenden Zustand. In der Mikroelektronik wird der in Abb. 16 dargestellte Transistor aus Silizium hergestellt. Der Isolator über dem p-Gebiet ist Siliziumdioxid. Deshalb spricht man von einem Metall-Oxid-Halbleiter (englisch «Semiconductor») Feldeffekttransistor oder abgekürzt von einem MOS-FET. Außer als MOS-FET läßt sich der Feldeffekttransistor auch noch auf eine etwas andere Weise realisieren — als sogenannter Sperrschicht-FET. Für die Mikroelektronik auf der Basis des Siliziums ist der Sperrschicht-FET aber nur von untergeordneter Bedeutung und braucht hier nicht weiter besprochen zu werden.

Der bisher beschriebene und in Abb. 16 schematisch dargestellte MOS-FET ist nur einer von mehreren möglichen. Man nennt ihn

n-Kanal-MOS-FET. Ähnlich wie beim bipolaren Transistor können hier n und p vertauscht werden, wodurch wir zum p-Kanal-MOS-FET gelangen, wie er in Abb. 17b und 17d zu sehen ist. Die in Abb. 17a, b gezeigten MOS-FETs arbeiten im sogenannten Anreicherungsbetrieb, d. h., durch Anlegen einer positiven Spannung am Gate wird der n-Kanal und durch Anlegen einer negativen Spannung der p-Kanal mit Ladungsträgern angereichert, und der Transistor wird leitend. Es ist aber auch der umgekehrte Fall möglich. Eine Anreicherung kann bereits ohne Gate-Spannung vorliegen (Abb. 17c, d). Dann führt eine negative Spannung im Falle des n-Kanal-MOS-FET (Abb. 17c) und eine positive im Falle des p-Kanal-MOS-FET (Abb. 17d) zu einer Verarmung an Ladungsträgern im Kanal und damit zur Sperrung des Stromflusses zwischen Source- und Drain-Elektrode. Oft werden p-Kanal-MOS-FETs im Anreicherungsbetrieb und n-Kanal-MOS-FETs im Verarmungsbetrieb verwendet.

MOS-FETs sind, ebenso wie bipolare Transistoren, elektrisch betätigbare Schalter für elektrische Ströme. Wir wollen uns auch hier wieder über die wichtigsten Kenngrößen dieser Schalter, vor allem über ihre Verzögerungszeit, Klarheit verschaffen. Nimmt man an, daß die für das Umschalten des MOS-FET erforderliche Spannungsänderung am Gate momentan, d. h. ohne Zeitverzug, entsteht, so ist die Schaltzeit durch die Zeitspanne bestimmt, die ein Ladungsträger zum Durchlaufen des Kanals benötigt. Das ist bei Kanallängen um $1\ \mu\text{m}$ eine extrem kurze Zeit im Bereich einiger 10^{-12} s. Aber

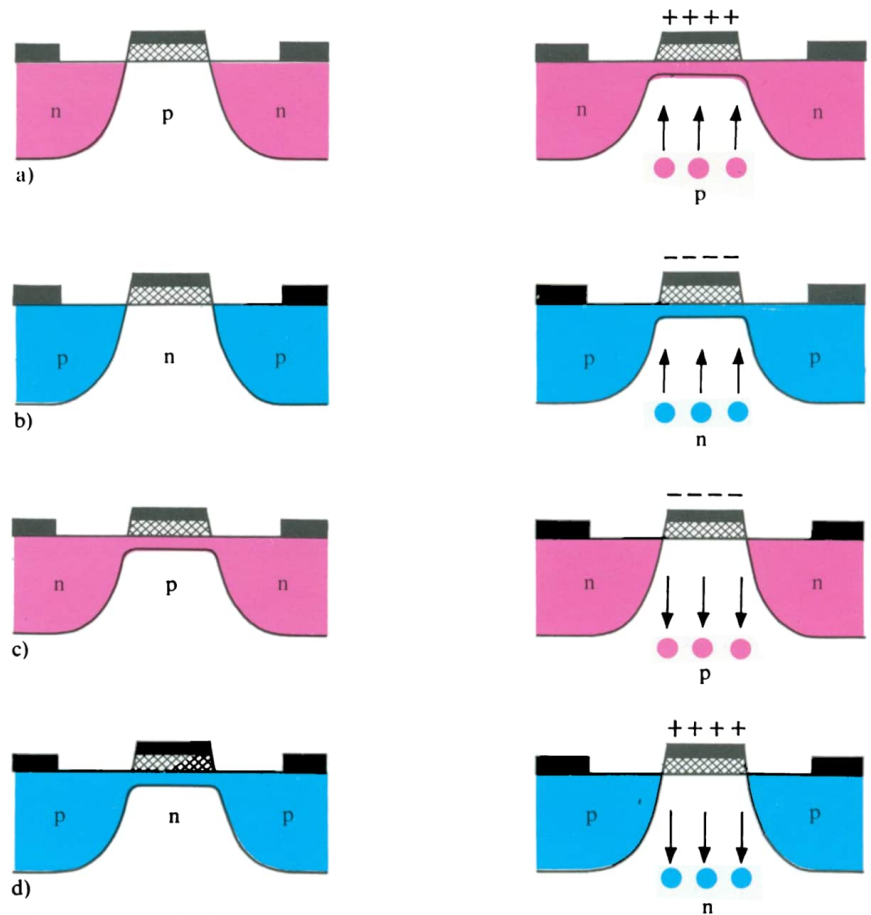


Abb. 17 Unterschiedliche Typen von MOS-FETs.
Links ohne Gate-Spannung — rechts mit Gate-Spannung

unsere Annahme über das momentane Entstehen der Schaltspannung am Gate ist unrealistisch. Sie läßt außer acht, daß dazu Ladungen auf die Gate-Elektrode fließen müssen. Die dafür erforderliche Zeit hängt von der Kapazität C des Gate und vom Widerstand R der Verbindungsleitung zum Gate ab. Sie ist das Produkt aus R und C . Nimmt man für $C \approx 10^{-12}$ Farad und für $R \approx 10^3$ Ohm, so erhält man für $RC = 10^{-9}$ s, also eine wesentlich längere Schaltzeit, als wir sie oben abgeschätzt hatten.

Meist ist sie die wirkliche Verzögerungszeit T des MOS-FET. Das gleiche gilt übrigens auch für den bipolaren Transistor. Wegen der relativ großen umzuladenden Kapazitäten und wegen der hohen Widerstände, über die das Umladen geschehen muß, sind MOS-FETs in der Regel langsamer als bipolare Transistoren. Da beim MOS-FET die Kapazität wesentlich durch die Gatefläche bestimmt ist und diese um so kleiner gemacht werden kann, je kürzer der Kanal ist, läuft die Forderung nach

geringen Verzögerungszeiten auf kleine Kanallängen hinaus. Wie verhält es sich mit der zum Schalten erforderlichen Leistung? Wie wir gesehen haben, wird der MOS-FET-Schalter über die Gate-Spannung betätigt. Die Schaltleistung ergibt sich als Produkt aus der Spannung, die am Gate anliegt, und dem Strom, der über das Gate abfließt. Während die Gate-Spannung etwa ebenso groß ist wie die Emitter-Basis-Spannung im Falle des bipolaren Transistors, ist der Gate-Strom außerhalb der Schaltakte wegen der isolierenden Schicht unter der Gate-Elektrode praktisch gleich 0. Während eines Schaltaktes ist er der Aufladestrom für die Gate-Elektrode. Dieser ist aber auch so gering, daß die zum Schalten benötigte Leistung beim MOS-FET dennoch um Größenordnungen kleiner bleibt als beim bipolaren Transistor. Wir haben bisher lediglich die Leistung betrachtet, die zum Schalten eines Transistors benötigt wird. Was bei einem Transistor geschaltet wird, ist aber ebenfalls eine elektrische Leistung, nämlich die im Emitter-Kollektor-Kreis des Bipolar-Transistors bzw. im Source-Drain-Kreis des MOS-FET umgesetzte eigentliche Verbrauchsleistung. Diese ist im allgemeinen wesentlich größer als die Schaltleistung. Ihr genauer Wert hängt von der angelegten Spannung und vor allem von der Art der Zusammenschaltung der Transistoren und der anderen Bauelemente in einer Schaltung ab. Auch die Verzögerungszeit wird durch die Zusammenschaltung der Transistoren und die Betriebsspannung wesentlich beeinflusst. Der Leistungsverbrauch und die Verzögerungszeit einer

Schaltung dürfen überhaupt nicht mit der Schaltleistung und der Verzögerungszeit eines einzelnen Transistors, für die wir oben Werte angegeben haben, verwechselt werden. Letztere sind Minimalwerte für die Verzögerungszeit und den Leistungsverbrauch einer Schaltung, die nicht unterschritten werden können. Sie geben uns eine Orientierung, wie gut wir die Eigenschaften des einzelnen Transistors in einer Schaltung ausnutzen bzw. wie gut eine Schaltung mit vorgegebenen Transistoren überhaupt sein kann. Im einzelnen werden wir auf die Güte mikroelektronischer Schaltungen in Kapitel 4 noch eingehen. Auf Grund unserer bisherigen Einsichten können wir aber bereits feststellen, daß für sehr schnelle Schaltungen bipolare Transistoren verwendet werden sollten. Braucht die Schaltung nicht extrem schnell zu sein, wird aber Wert auf geringen Leistungsverbrauch gelegt, so ist den MOS-FETs der Vorzug zu geben.

Widerstände. Kondensatoren

Mit den Transistoren und Halbleiterdioden haben wir wichtige Bauelemente kennengelernt, wie sie in mikroelektronischen Schaltkreisen verwendet werden. Aber bekanntlich braucht man zum Aufbau einer elektronischen Schaltung im allgemeinen auch noch Widerstände, Kondensatoren (Kapazitäten) und Spulen (Induktivitäten). Die Wirkungsweise von Widerständen und Kondensatoren ist wesentlich einfacher als die von Transistoren und Dioden. Wir brauchen darauf hier nicht weiter einzugehen. Nicht ganz so einfach wie die Wirkungsweise ist aber die Realisierung dieser Bauelemente in der Mikroelektronik. Die

heute am häufigsten genutzte Realisierungsmöglichkeit besteht im Ersatz von Widerständen und Kondensatoren durch Transistoren. Legt man z. B. die Basis eines npn-Transistors ein für allemal an eine feste negative Spannung an, so stellt die Emitter-Kollektor-Strecke einen relativ hohen Widerstand dar. Bei einem MOS-FET kann man das Gate und die ihm gegenüberliegende sog. Substratelektrode als die leitenden Platten und die Oxidschicht unter dem Gate als isolierende Zwischenschicht eines Kondensators benutzen und auf diesem Wege eine Kapazität bestimmter Größe realisieren. Daneben wird in der Mikroelektronik aber auch die direkte Herstellung von Widerständen und Kondensatoren angewandt. Einen Widerstand kann man z. B. dadurch realisieren, daß man in einem n-Gebiet einen p-leitenden Streifen erzeugt. Durch die Wahl der geometrischen Abmessungen des p-Streifens läßt sich sein Widerstand innerhalb bestimmter Grenzen einstellen, vorausgesetzt man sorgt durch Anlegen einer Sperrspannung zwischen n- und p-Gebiet dafür, daß der Strom nur über den p-Streifen fließen kann. Um einen Kondensator herzustellen, kann man auf die p-dotierten Source- und Drain-Gebiete eines n-Kanal MOS-FET überhaupt verzichten und lediglich mit dem Gate, der darunterliegenden Oxidschicht und dem n-Silizium unter der Oxidschicht arbeiten. Wir verweisen auf Kapitel 5, wo die hier angedeuteten Realisierungen von Widerständen und Kondensatoren in Abb. 38 dargestellt sind. Zur Realisierung von Induktivitäten in der Mikroelektronik ist anzumerken,

daß sie sich als Einzelbauelemente überhaupt nicht herstellen lassen. Ihre Funktion muß durch Zusammenschalten der anderen Bauelemente simuliert werden, oder Induktivitäten müssen bei der Konzipierung des Schaltkreises überhaupt vermieden und außen dazugeschaltet werden.

Optische Anzeigebaulemente.

Halbleitersensoren

Die Verarbeitung der Information in einer elektronischen Schaltung ist in vielen Fällen noch nicht das ganze Informationsproblem, wie es in der Praxis vorkommt. Die in der Schaltung zu verarbeitende Information muß oft erst durch Messen gewonnen werden, und das Ergebnis der Informationsverarbeitung muß optisch angezeigt werden. Für beide Zwecke, für die Informationsgewinnung und für die optische Informationsanzeige, stehen Halbleiterbauelemente zur Verfügung. Ihr Vorteil gegenüber anderen Lösungen ist ihre gute Angepaßtheit an die mikroelektronischen Schaltkreise, die ebenfalls aus Halbleiterbauelementen aufgebaut sind. Dazu kommt, daß sie dank der hocheffektiven Verfahren der Halbleitertechnologie billig hergestellt werden können.

Für die optische Anzeige von Symbolen und Zeichen kommen neben Bauelementen auf der Basis von Flüssigkristallen auch Halbleiterbauelemente zum Einsatz. Die sogenannten ac-Dünnschicht-Elektrolumineszenz-Displays aus Mangan-dotiertem Zinksulfid dienen der Anzeige größerer Informationsinhalte. Für kleinere Informationsmengen sind Anzeigen aus Licht-Emitterdioden, sogenannte LEDs, geeignet. Bei den LEDs wird die strahlende

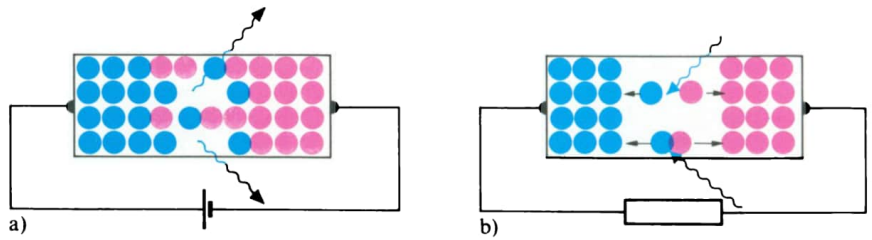


Abb. 18 Prinzip einer Licht-Emitter-Diode (LED) (a) und einer Photodiode (b)

Rekombination von Elektronen und Löchern ausgenutzt. Ein Elektron im Leitungsband bzw. in einem Störniveau rekombiniert mit einem Loch im Valenzband bzw. in einem Störniveau und emittiert dabei ein Photon. Damit sich dieser Vorgang kontinuierlich wiederholen kann, müssen ständig Elektronen und Löcher nachgeliefert werden. Das geschieht bei den LEDs durch das Anlegen einer Spannung an einen in Durchlaßrichtung gepolten pn-Übergang (Abb. 18). Dadurch werden Überschuß-Elektronen in das p-Gebiet und Überschuß-Löcher in das n-Gebiet injiziert, die mit den dort vorhandenen Majoritätsladungsträgern unter Emission von Licht rekombinieren. Die elektrische Energie der Spannungsquelle wird also direkt in Licht umgewandelt.

Bestrahlt man umgekehrt einen pn-Übergang mit Licht, so werden Elektron-Loch-Paare erzeugt, die dann im elektrischen Feld des pn-Übergangs getrennt werden (Abb. 18b). Durch den pn-Übergang fließt ein elektrischer Strom, dessen Größe ein Maß für die Intensität des Lichtes ist. Man bezeichnet ein solches Bauelement als *Photodiode*. Eine Photodiode wandelt ein Lichtsignal in ein elektrisches Signal um, das in einem mikroelektronischen Schaltkreis direkt verarbeitet werden kann.

Die Photodiode ist ein Beispiel für ein Halbleiterbauelement zur Informationsgewinnung für die Mikroelektronik. Mitunter bezeichnet man Halbleiterbauelemente, die nichtelektrische Signale in elektrische umwandeln, als *Halbleitersensoren*. Derartige Sensoren gibt es heute außer für Licht bereits für die Messung von Drücken, Temperaturen oder Konzentrationen chemischer Stoffe. Die Funktionsprinzipien dieser Sensoren lassen sich auf der Grundlage der oben behandelten Halbleitereffekte leicht verstehen. Wir wollen aber auf eine detaillierte Darstellung verzichten. Die Produktion und der Einsatz von Halbleitersensoren befinden sich gegenwärtig mit Ausnahme der optischen noch in den Anfängen. Die Entwicklung schreitet aber gerade auf diesem Gebiet schnell voran, handelt es sich bei den Sensoren doch um Bauelemente, ohne die in vielen Fällen die Anwendung der Mikroelektronik nicht möglich ist.

Nachdem wir die Funktionsweise der für die Elektronik wichtigen Halbleiterbauelemente, darunter vor allem die der Transistoren, kennengelernt haben, werden wir uns im nächsten Kapitel damit beschäftigen, wie durch das Zusammenschalten dieser Bauelemente, also durch elektronische Schaltungen, Information verarbeitet werden kann.

4. Informationsverarbeitung mit Hilfe elektronischer Schaltungen

Information bedarf eines materiellen Trägers. In der Elektronik ist dieser Träger die Elektrizität. Informationen lassen sich durch unterschiedlich starke elektrische Ströme oder Spannungen darstellen, ähnlich wie bei einer Fotografie die Bildinformation durch unterschiedlich starke Schwärzungen des Fotopapiers dargestellt ist. Die örtlich variierenden Helligkeitswerte einer Fotografie können sogar direkt in ein elektrisches Spannungssignal, d. h. in zeitlich variierende Spannungswerte umgewandelt werden, indem man die Fotografie mit einem Fotodetektor abtastet. Ähnlich wie auf dem Fotopapier alle möglichen Helligkeitswerte vorkommen — innerhalb gewisser Grenzen selbstverständlich — so sind im elektrischen Spannungssignal alle möglichen Spannungswerte vertreten. Große Helligkeitswerte entsprechen großen Spannungen, kleine Helligkeitswerte kleinen Spannungen.

Analoge und digitale Informationsverarbeitung

Die Information in Form des Spannungssignals ist der ursprünglichen Information auf dem Fotopapier *analog*. Man spricht immer dann von *analoger Darstellung* der Information, wenn die ursprüngliche Information im obigen Sinne wertgetreu auf die elektrische Spannung übertragen wird. Die Verarbeitung von Information in Form

analoger Spannungssignale nennt man *analoge Informationsverarbeitung*. Sie wird z. B. in Radio- und Fernsehgeräten angewandt. Daneben gibt es die *digitale Informationsverarbeitung*, die in der Mikroelektronik ganz besondere Bedeutung besitzt.

Um zu verstehen, worum es bei der digitalen Informationsverarbeitung geht, zerlegen wir das analoge Spannungssignal in Abb. 19 in einzelne Abschnitte, indem wir die Zeitachse in kleine, gleichgroße Intervalle aufteilen. Genauso wie in der Musik bezeichnet man diese Zeitintervalle als *Takte*. In jedem dieser Takte kann die Spannung annähernd als kon-

stant angesehen werden, vorausgesetzt, wir haben die Taktlänge klein genug gewählt. Eine bestimmte Spannung, die genau einen Takt lang andauert, wird als *Spannungsimpuls* bezeichnet. Das gesamte Spannungssignal in Abb. 19 kann dann näherungsweise als lückenlose zeitliche Aufeinanderfolge von Spannungsimpulsen unterschiedlicher Größe aufgefaßt werden. Die analoge Informationsverarbeitung operiert mit Impulsfolgen aller möglichen Spannungswerte. Bei der digitalen Informationsverarbeitung werden nur zwei Spannungswerte benutzt, ein minimaler, der mit 0 bezeichnet wird, und ein maximaler, der mit 1

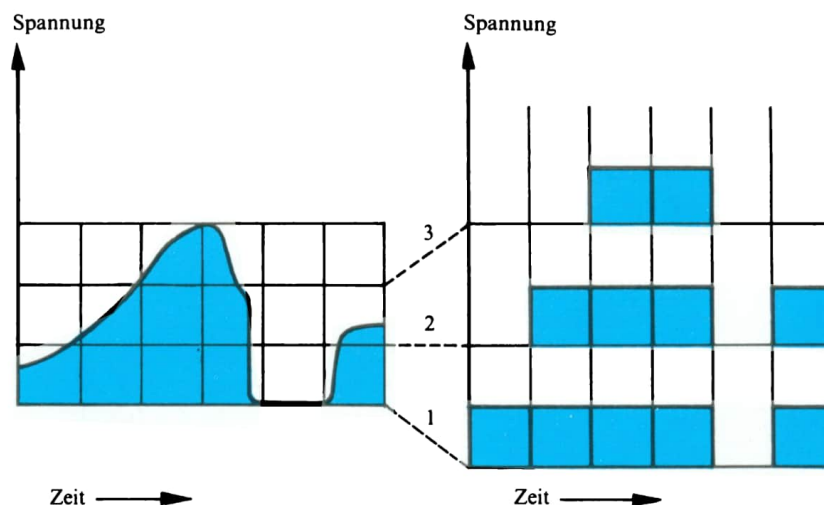


Abb. 19 Analoge und digitale Impulsfolgen und ihre Umformung ineinander

bezeichnet wird. Jede analoge Impulsfolge läßt sich näherungsweise in eine größere Anzahl von digitalen Impulsfolgen auflösen, die insgesamt annähernd die gleiche Information verkörpern wie die analoge Impulsfolge. Ein Verfahren zur Umwandlung einer analogen in eine digitale Impulsfolge ist in Abb. 19 veranschaulicht. Man teilt die Spannungsskala in eine bestimmte Anzahl n gleicher Spannungsintervalle ein ($n = 3$ in Abb. 19) und stellt für jedes Intervall eine eigene elektrische Leitung zur Verfügung — für das erste Intervall Leitung 1, für das zweite Intervall Leitung 2 usw. Fassen wir den ersten Takt in Abb. 19 ins Auge. Dort ist das erste Spannungsintervall vom Impuls überdeckt, das zweite und dritte Intervall ist leer. Man gibt deshalb in die Leitung 1 einen 1-Impuls und in die beiden anderen Leitungen 0-Impulse. Im zweiten Takt ist das erste Intervall vollständig, das zweite teilweise und das dritte nicht überdeckt. Man gibt also einen 1-Impuls in die erste, einen 1-Impuls in die zweite und einen 0-Impuls in die dritte Leitung. Das wiederholt man für alle Zeitakte. Auf diese Weise entstehen n Impulsfolgen, die nur aus 0- und 1-Impulsen zusammengesetzt sind. Die analoge Impulsfolge wird durch die digitalen Folgen allerdings nur in vergrößerter Form wiedergegeben, weil die Spannung durch n diskrete Werte beschrieben wurde. Die Rundungsfehler sind um so kleiner, je feiner das Maschennetz ist, das wir über die ursprüngliche analoge Impulsfolge legen. Mit der Feinheit der Maschen wächst aber auch die Anzahl der verschiedenen Impulsfolgen und die Anzahl der Impulse in jeder

einzelnen Folge, die pro Sekunde zu verarbeiten sind. Höhere Genauigkeit erfordert also höheren digitalen Verarbeitungsaufwand. Um bei der analogen Informationsverarbeitung eine hohe Genauigkeit zu erzielen, benötigt man Bauelemente von hoher Qualität. Sie müssen insbesondere eine sehr gute Linearität besitzen, d. h., die Spannung am Ausgang muß sich genau linear mit der Spannung am Eingang des Bauelementes ändern, sonst wird die Information am Ausgang gegenüber der am Eingang verfälscht. Die digitale Form der Informationsverarbeitung ist im allgemeinen zuverlässiger und weniger störanfällig gegenüber den Unvollkommenheiten der einzelnen Bauelemente. Sie benötigt dafür aber wesentlich mehr einzelne Bauelemente als die analoge oder ist, wenn nicht genügend Bauelemente zur Verfügung stehen, langsamer als die analoge. Mit Hilfe der digitalen Informationsverarbeitung können eine höhere Zuverlässigkeit und eine größere Toleranz gegenüber den Ungenauigkeiten der einzelnen Bauelemente um den Preis einer erhöhten Anzahl von Bauelementen erkaufte werden. Diesen Preis zu zahlen fällt der Mikroelektronik nicht besonders schwer. Sehr viele, nicht übermäßig genaue, aber kleine, schnelle und billige Bauelemente auf kleinstem Raum unterzubringen, ist genau das, wozu die Mikroelektronik imstande ist. Digitale Informationsverarbeitung und Mikroelektronik haben sich in gewisser Weise gesucht und gefunden. Mit Hilfe der Mikroelektronik kann die hohe Qualität weniger präziser Verarbeitungsschritte durch die große Quantität vieler schneller, nicht so präziser Schritte

ersetzt werden. Auch störende Einflüsse aus der Umwelt oder aus der Verarbeitungsanlage selbst können bei der digitalen Informationsverarbeitung besser verkraftet werden als bei der analogen.

Binärzahlen

Wir haben gesehen, auf welche Weise analoge Impulsfolgen in digitale umgewandelt werden können. Oft liegt die Information primär in Form von *Dezimalzahlen* vor. Damit die Verarbeitung auch in diesen Fällen digital erfolgen kann, benötigt man ein Verfahren, mit dem Dezimalzahlen in Binärzahlen umgewandelt werden können und umgekehrt, Binärzahlen in Dezimalzahlen. Dieses Verfahren soll nun kurz dargestellt werden. Wir erinnern zunächst daran, daß im Dezimalsystem die n -te Stelle einen Stellenwert 10^{n-1} besitzt, d. h. die erste Stelle $10^0 = 1$ (Einer), die zweite Stelle $10^1 = 10$ (Zehner), die dritte Stelle $10^2 = 100$ (Hunderter) usw. Die vierstellige Dezimalzahl 6347 bedeutet ausführlich geschrieben $6 \cdot 10^3 + 3 \cdot 10^2 + 4 \cdot 10^1 + 7 \cdot 10^0$. Die beiden einstelligen Binärzahlen sind 0 und 1. Das Binärsystem ist analog dem Dezimalsystem aufgebaut, nur daß den einzelnen Stellen nicht verschiedene Potenzen von 10, sondern von 2 entsprechen. Die n -te Stelle hat den Stellenwert 2^{n-1} . Die vierstellige Binärzahl 0101 bedeutet also $0 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 5$. Mit einer vierstelligen Binärzahl läßt sich maximal die Zahl $1111 = 2^3 + 2^2 + 2^1 + 2^0 = 15$ darstellen. Um bis $2^{10} - 1 = 1023$ zählen zu können, benötigt man eine 10stellige Binärzahl, zum Zählen bis $2^{20} - 1 = 1\,048\,575$ eine 20stellige. Jede Zahl kann sowohl im Dezimal-

system als auch im Binärsystem dargestellt werden. Im Binärsystem ist die Anzahl der Stellen zwar größer als im Dezimalsystem, aber die Anzahl von Zeichen für jede Stelle ist wesentlich kleiner, nämlich 2 statt 10 im Falle des Dezimalsystems. Das Binärsystem hat gegenüber dem Dezimalsystem den Vorteil, daß zur Darstellung der Zahlen ein geringerer Vorrat an Zeichen erforderlich ist. Davon überzeugt man sich leicht durch folgende Abschätzung. Für die Darstellung aller Zahlen unter 999999 benötigt man $6 \cdot 10 = 60$ Zeichen im Dezimalsystem, aber nur $20 \cdot 2 = 40$ im Binärsystem. Trotzdem ist das Binärsystem noch nicht das mit der niedrigsten Zeichenzahl. Das Minimum an Zeichen wird erreicht, wenn die Stellenwerte Potenzen der Zahl $e = 2,718$ sind. Aber man legt natürlich Wert darauf, daß die Stellen Potenzen einer ganzen Zahl darstellen. Die bestmöglichen Kandidaten dafür sind «zwei» und «drei». Nicht zuletzt aus Gründen der besseren Realisierbarkeit in elektronischen Schaltungen wählt man das Binärsystem statt des «Ternärsystems». Trotz der informationstheoretischen Überlegenheit des Binärsystems gegenüber dem Dezimalsystem — die übrigens von G. W. Leibniz bereits 1679 bemerkt wurde — löst man sich beim maschinellen Rechnen aber nicht gänzlich vom Dezimalsystem. Der Grund dafür ist letztlich, daß der Mensch zehn Finger hat und sich das Gedächtnis im Laufe der Entwicklung des Menschen an das Dezimalsystem angepaßt hat. Man verwendet deshalb das Dezimalsystem überall dort, wo der Kontakt zum Menschen herge-

stellt werden muß, also insbesondere bei der Ein- und Ausgabe der Daten. Innerhalb der Informationsverarbeitungsanlage wird jedoch mit Binärzahlen gerechnet. Dabei wird z. T. die sogenannte BCD-Darstellung, d. h. die Binär-Codierung von Dezimalzahlen, verwendet. Der Stellenwert der Dezimalziffern aus dem Dezimalsystem bleibt dabei erhalten. Für jede Dezimalziffer benötigt man, wie bereits erwähnt, eine vierstellige Binärzahl, nämlich:

0 = 0000; 1 = 0001; 2 = 0010;
 3 = 0011; 4 = 0100; 5 = 0101;
 6 = 0110; 7 = 0111; 8 = 1000;
 9 = 1001.

Die Möglichkeiten der vier Binärstellen werden bei der BCD-Darstellung zwar nicht voll genutzt (man kann mit vier Stellen alle Zahlen bis zur 15 darstellen), aber die Beziehung zwischen dezimaler und dualer Darstellung ist überschaubarer als bei der konsequent binären Darstellung. Die Zahl 2536 z. B. lautet in der BCD-Darstellung 0010 0101 0011 0110. Sie ist also aus Gruppen von jeweils vier Binärzahlen aufgebaut, die man auch Tetraden nennt. In der rein binären Darstellung ist $2536 = 100111101000$. Damit haben wir das Wichtigste über die Darstellung von Dezimalzahlen im Binärsystem kennengelernt. Es ist jetzt nur noch ein kleiner Schritt, um auch Buchstaben, mathematische Formelzeichen oder ganz allgemein *Symbole* und geordnete Gruppen aus solchen Symbolen wie Worte oder Sätze in Form von Binärzahlen zu codieren. Um das lateinische Alphabet darzustellen, kann man z. B. jedem der 26 Buchstaben eine Zahl zwischen 0 und 25

zuordnen, also etwa dem a die 0, dem b die 1, dem c die 2 usw. und diese Zahlen anschließend als fünfstellige Binärzahl verschlüsseln. Ein Buchstabe repräsentiert dann eine Informationsmenge von 5 bit — in Wahrheit sind es nur $\log_2 26 = 4,7$, was daran liegt, daß wir mit dem Alphabet die Möglichkeiten einer fünfstelligen Binärzahl nicht ausschöpfen. Was uns hier am Beispiel der binären Darstellung des Alphabets begegnet und was sich bereits oben bei der BCD-Darstellung von Dezimalzahlen andeutete, gilt ganz allgemein für die binäre Darstellung von Information. Die zur Darstellung benötigte Anzahl von Binärzeichen ist nicht unbedingt gleich dem in bit ausgedrückten Informationsgehalt, sondern der Informationsgehalt kann kleiner sein als die Zahl der Binärzeichen. Für letztere gibt es deshalb eine eigene Maßeinheit — das *Bit*. Im Unterschied zum *bit*, das wir als Maßeinheit für die Informationsmenge bereits kennen, wird es groß geschrieben. Das *Bit* ist die Maßeinheit für die Länge einer Binärzahl. Eine n-stellige Binärzahl besitzt die Länge n Bit. Ob sie auch n bit Information enthält, ist eine ganz andere Frage. Die Maßzahl beim «Bit» kann immer nur eine ganze Zahl sein, halbe oder zehntel Bits gibt es nicht, weil es keine halben oder zehntel Binärzeichen gibt. Die Informationsmenge kann aber durchaus 0,5 bit oder 0,1 bit betragen. Es stimmt also nicht, wenn mitunter behauptet wird, das «bit» sei die kleinste Informationsmenge, sozusagen das Elementarquantum der Information. Wohl aber ist 1 Bit die kleinste Länge einer Binärzahl. Bei 8 Bit

spricht man von 1 Byte. 2^{10} Bit bezeichnet man als KBit, 2^{20} als MBit. Da $2^{10} = 1024$ und $2^{20} = 1\,048\,576$, entspricht 1 KBit etwa 1 Tausend Bit und 1 MBit etwa 1 Million Bit. Je kleiner die Bit-Zahl ist, die zur Darstellung einer bestimmten Information benötigt wird, umso größer ist die Informationsdichte der Darstellung. Das Ideal ist 1 bit/Bit, das aber nur selten erreicht wird. Bei der binären Darstellung des Alphabets z. B. beträgt die Informationsdichte 4,7/5 bit/Bit = 0,9 bit/Bit, bei der BCD-Darstellung einer 4stelligen Dezimalzahl sind es $\log_2 10^4/16$ bit/Bit = 0,8 bit/Bit. Um alle Zeichen, die zum Schreiben eines Textes benötigt werden, also auch Großbuchstaben, Ziffern, Satzzeichen, Trennungslücke und andere Steuerzeichen, binär darstellen zu können, braucht man 7 Bit (ASCII-Code).

Nach dieser Abschweifung über die Maßeinheit Bit, die uns später von Nutzen sein wird, kehren wir zu den Binärzahlen zurück. Das Rechnen mit mehrstelligen Binärzahlen kann nach dem gleichen Verfahren durchgeführt werden wie das mit mehrstelligen Dezimalzahlen. Als Beispiel betrachten wir die Addition von zwei vierstelligen Binärzahlen: 0011 und 0110. Wir schreiben sie nach Adam Ries untereinander:

0011
0110

1001

und addieren stellenweise. Dabei tritt ein Übertrag 1 für die nächste Stelle auf, wenn in der vorhergehenden Stelle 1 und 1 addiert werden mußten. Nach diesem Schema haben wir als Summe die Zahl 1001 erhalten. Wir überprüfen im Dezimal-

system, ob wir richtig gerechnet haben:
0011 = 3, 0110 = 6, 1001 = 9.
Das Ergebnis stimmt also.
Vergegenwärtigen wir uns noch einmal unser Vorgehen bei der Addition. Wir haben sie in mehreren gleichartigen Schritten durchgeführt. In jedem einzelnen Schritt wurden zwei einstellige Binärzahlen addiert. Wir können dafür auch sagen, daß in jedem einzelnen Schritt zwei einstelligen Binärzahlen eine dritte (maximal zweistellige) Zahl zugeordnet wurde — dem Paar (0,0) die 0, dem Paar (0,1) oder (1,0) die 1 und dem Paar (1,1) die 1 0. Auf solche Zuordnungen zwischen Binärzahlen läßt sich nicht nur die Addition zurückführen, sondern auf diese Weise können auch alle anderen arithmetischen Rechenoperationen, also die Subtraktion, Multiplikation und Division, ja überhaupt alle Probleme der Informationsverarbeitung gelöst werden. Die Zuordnungsvorschriften, wie wir sie bei der Addition kennengelernt haben, sind nur noch nicht elementar genug. Sie lassen sich

Elementare logische Operationen
auf noch elementarere, sogenannte *Boolesche* Zuordnungsvorschriften, die nur mit einstelligen Binärzahlen operieren, zurückführen. Zuordnungsvorschriften bezeichnet man in der Mathematik auch als Funktionen. Wir werden deshalb im folgenden statt von einer Zuordnungsvorschrift von einer Funktion sprechen. Funktionen hängen im allgemeinen von mehreren Größen $x, y, \dots$ ab, für die man Werte in die Funktion einsetzen kann. Außerdem haben Funktionen einen bestimmten Vorrat an Funktionswerten z , die den Eingangswerten zugeordnet

sind. Man spricht von ein- oder mehrwertigen Funktionen, je nachdem, ob die Funktion von einer oder mehreren Größen abhängt. Für die Informationsverarbeitung kommt man mit ein- und zweiwertigen Booleschen Funktionen aus. Bei den einwertigen Funktionen sind zwei Eingangswerte, nämlich 0 und 1, und zwei Funktionswerte, ebenfalls 0 und 1, möglich. Es gibt dementsprechend vier verschiedene einwertige Boolesche Funktionen. Davon ist nur eine, und zwar die Negation, interessant, die anderen sind trivial. Durch die Negation wird der 0 die 1 und der 1 die 0 zugeordnet. Man schreibt die Negation in der Form $z = \bar{x}$ und meint damit, daß sich $z = 1$ für $x = 0$ und $z = 0$ für $x = 1$ ergibt (siehe Tabelle 2a).

Tab. 2a Negation

x	$\bar{x}$
0	1
1	0

Der Begriff «Negation» ist dem philosophisch vorgebildeten Leser von Hegels «Negation der Negation» her bekannt. In diesem philosophischen Zusammenhang wird die «Negation» in mehr qualitativem Sinne gebraucht: Das Alte wird doppelt negiert. Was dabei herauskommt ist und ist auch nicht das Alte — es ist das Alte auf einer höheren Stufe. Die doppelte Anwendung der Negation im Sinne der Logik ergibt genau wieder die Ausgangsgröße. Die logische Negation und ihre abermalige Negation ist auch Grundlage für ein bekanntes Rätsel: Von einem Menschen ist be-

kennt, daß er entweder stets die Wahrheit sagt oder stets die Unwahrheit. Es ist aber nicht bekannt, ob das eine oder das andere zutrifft. Wie muß man diesen Menschen an einer Weggabelung fragen, um in jedem Falle die richtige Antwort auf die Frage nach dem rechten Weg zu bekommen? Man muß so fragen, daß er zu einer Aussage über eine seiner Aussagen veranlaßt wird, z. B.: Was hätten Sie einem Wanderer gestern geantwortet, wenn er Sie nach dem rechten Weg gefragt hätte? Ist der Mensch immer ehrlich, ist jeder Teil der Antwort richtig, insbesondere auch die den Weg betreffende. Lügt der Mensch immer, so sagt er auch nicht die Wahrheit über das, was er gestern einem Wanderer geantwortet hätte. Er verkehrt die richtige Antwort über den Weg also zweimal in ihr Gegenteil, und dadurch wird die Antwort richtig. Die Negation ist eine einwertige Boolesche Funktion. Für die Informationsverarbeitung benötigt man, wie bereits erwähnt, auch zweiwertige Funktionen. Bei ihnen sind vier verschiedene Eingangswerte, nämlich (0,0), (0,1), (1,0), (1,1), und zwei verschiedene Funktionswerte, nämlich 0 und 1, möglich, insgesamt gibt es also $2^4 = 16$ verschiedene zweiwertige Boolesche Funktionen. Man kommt aber mit einer kleinen Auswahl unter ihnen aus, auf die sich die übrigen Funktionen zurückführen lassen. Zu dieser Auswahl gehört die Negation, die wir bereits als einstellige Funktion kennengelernt haben, die aber auch unter den zweistelligen Funktionen vorkommt, nämlich als eine Funktion, bei der der Funktionswert nur von einer der beiden Größen, etwa x , abhängt und

den Wert $z = \bar{x}$ hat. Ferner gehören dazu die Disjunktion und die Konjunktion. Bei der Disjunktion wird den beiden Größen x und y gemäß Tabelle 2b eine Größe $z = x \vee y$ zugeordnet:

Tab. 2b Disjunktion (Oder)

x	y	$x \vee y$
0	0	0
0	1	1
1	0	1
1	1	1

Die Disjunktion ist die logische Operation «oder»: wenn x nicht ist und y nicht, dann ist « x oder y » nicht. Wenn $x(y)$ ist, $y(x)$ nicht ist, dann ist « x oder y ». Wenn x ist und y ist, dann ist ebenfalls « x oder y ». Die Konjunktion ordnet den Elementen x und y gemäß Tabelle 2c das Element $z = x \wedge y$ zu:

Tab. 2c Konjunktion (Und)

x	y	$x \wedge y$
0	0	0
0	1	0
1	0	0
1	1	1

Die Konjunktion ist die logische Operation «und»: Wenn x nicht ist und y nicht ist, dann ist « x und y » nicht, wenn $x(y)$ und $y(x)$ nicht ist, dann ist « x und y » nicht, wenn x ist und y ist, dann ist « x und y ». Die logischen Operationen «oder» und «und» sind völlig eindeutig. Für den normalen Sprachgebrauch

des Wortes «oder» gilt das nicht. Ist x z. B. die Aussage «er ist zu Hause» und y die Aussage «sie ist zu Hause», so ist die Disjunktion von x , y , $x \vee y$ die Aussage «er ist zu Hause oder sie ist zu Hause». Formal logisch bedeutet das, daß entweder er allein zu Hause ist oder sie allein zu Hause ist, oder daß beide zu Hause sind. Umgangssprachlich würde man darunter aber eher verstehen, daß entweder er oder sie zu Hause ist, obwohl auch eine Auslegung im Sinne der formal logischen Bedeutung denkbar wäre. Umgangssprachlich hat «oder» also offensichtlich zwei Bedeutungen — eine ausschließende, im Sinne von «entweder oder» und eine verbindende, die der formal logischen entspricht. Die Sprache benötigt diese Zweideutigkeiten. Sie ermöglichen eine größere Variabilität des Ausdrucks. Eindeutigkeit wird durch den sprachlichen Gesamtzusammenhang erreicht. Mit der umgangssprachlichen Bedeutung von «und» ist es anders. Unter der Aussage «er ist zu Hause und sie ist zu Hause» kann nur verstanden werden, daß sowohl er als auch sie zu Hause ist, so wie es der logischen Bedeutung von «und» entspricht. Um die umgangssprachlichen Zweideutigkeiten zu vermeiden, wählt man in der Mikroelektronik lieber die englischen Bezeichnungen «OR» und «AND», wenn man die logischen «oder» und «und» meint. Es ist nützlich, für die Funktionen Negation, Disjunktion und Konjunktion graphische Symbole einzuführen. Diese sind in Abb. 20 dargestellt. Auf die Operationen «Negation», «Disjunktion» und «Konjunktion» läßt sich jedes Problem der digitalen

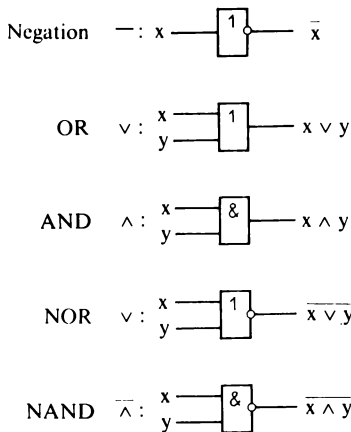


Abb. 20 Schaltsymbole für die elementaren logischen Operationen

Informationsverarbeitung reduzieren. Man bezeichnet sie deshalb als die *elementaren* logischen Operationen. Die Vereinfachung läßt sich sogar noch weiter treiben. Man benötigt überhaupt nicht alle drei genannten Operationen, sondern kommt mit einer einzigen Operation aus, und zwar entweder mit einer sogenannten NOR-Operation, die durch Negation des Ergebnisses der OR-Verknüpfung entsteht:

$$x \text{ NOR } y = \overline{x \vee y},$$

oder mit der sogenannten NAND-Operation, die die Negation des Ergebnisses der AND-Verknüpfung bedeutet:

$$x \text{ NAND } y = \overline{x \wedge y}.$$

Aus praktischen Gründen verwendet man in der Mikroelektronik aber sowohl die NAND- als auch die NOR-Operation, und neben beiden auch noch OR, AND und die Negation.

Schaltungen für die elementaren logischen Operationen

Im folgenden soll besprochen werden, wie sich Schaltungen für diese elementaren logischen Operationen mit Hilfe von Transistoren und Widerständen aufbauen lassen. Wir gehen dabei schrittweise vor und betrachten zunächst eine elektromechanische Lösung, bevor wir auf die rein elektronische zu sprechen kommen. Wir wissen, daß Binärzahlen 0 und 1 durch das Vorhandensein einer minimalen oder maximalen Spannung dargestellt werden können. Ebenso gut lassen sie sich aber auch mittels eines mechanischen Schalters in einem Stromkreis realisieren: gesperrter Schalter ist 0, leitender Schalter 1. Die OR- und AND-Verknüpfung dieser Schalterstellungen ist mit Hilfe der in Abb. 21 dargestellten Schaltungen möglich. Das Ergebnis der Verknüpfung ist in beiden Fällen die Span-

nung, die zwischen dem mit A und dem mit T bezeichneten Anschluß anliegt. Keine Spannung heißt 0, Spannung vorhanden heißt 1. Wir betrachten zunächst die OR-Schaltung (Abb. 21 a). Sind beide Schalter in Abb. 21 a «aus», so befindet sich A auf dem Potential der unteren Elektrode. Die Spannung zwischen A und T ist also 0. Ist der linke Schalter «ein», der rechte «aus», so fließt ein Strom durch den linken Zweig und den Widerstand. Nehmen wir an, daß der eingeschaltete Schalter so gut wie keinen Widerstand hat, fällt die gesamte Spannung zwischen H und T am Widerstand R ab. Der obere Eingang des Widerstandes und damit auch der Ausgang A sind auf dem Potential H. Zwischen A und T besteht also eine Spannung, d. h., der Ausgang gibt 1 aus. Dasselbe geschieht, wenn statt des linken der rechte Schalter «ein» ist. Auch wenn beide

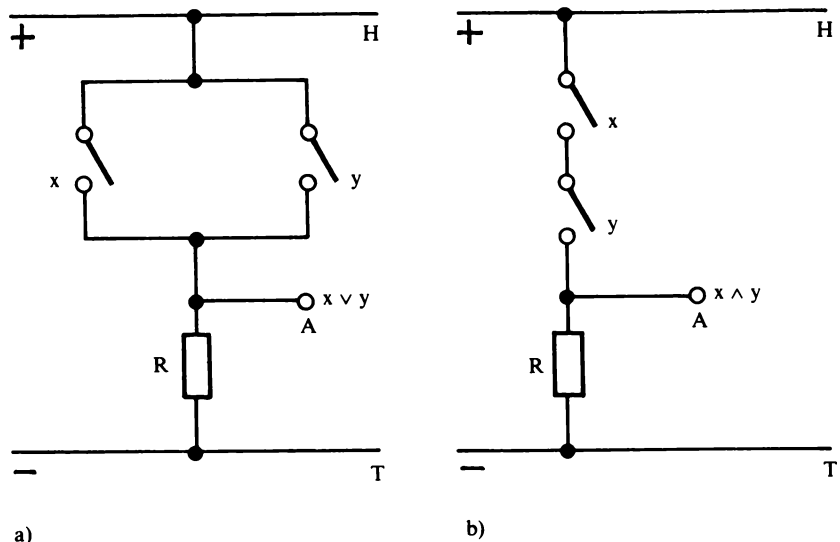


Abb. 21 Schaltungen für die OR- (a) und AND-Operationen (b) unter Verwendung mechanischer Schalter

Schalter «ein» sind, passiert nichts anderes als das Fließen eines Stromes durch den Widerstand R. Er ist ebenso groß wie in den Fällen, wo nur *ein* Schalter geöffnet war, er teilt sich im verzweigten Teil der Schaltung nur auf beide Zweige gleichmäßig auf. Ein Vergleich der binären Zahlenwerte am Ausgang mit der Wertetafel für OR in Tabelle 2b zeigt vollkommene Übereinstimmung. Am Ausgang A entsteht also tatsächlich das Ergebnis der OR-Verknüpfung der beiden Eingangsinformationen x, y. In gleicher Weise kann man sich klarmachen, daß am Ausgang A der Schaltung in Abb. 21 b das Ergebnis der AND-Verknüpfung der beiden Eingangsinformationen entsteht. Die detaillierte Ausführung dieser Überlegung überlassen wir dem Leser.

Auch die Negation kann mit einem mechanischen Schalter realisiert werden. In der Elektronik braucht man dafür einen Transistor. Die Schaltsymbole für bipolare und Feldeffekttransistoren sind in Abb. 22 dargestellt. Der Emitter wird durch einen Pfeil gekennzeichnet. Beim npn-Transistor liegt der Emitter normalerweise auf niedrigem Potential, der Kollektor auf hohem. Der Strom fließt vom hohen zum niedrigen Potential, also aus dem Emitter in der durch den Pfeil bezeichneten Richtung heraus. Beim pnp-Transistor liegt der Emitter auf hohem Potential, der Kollektor auf niedrigem, und der Strom fließt in der angegebenen Weise in den Emitter hinein. Das Potential an der Basis kann hoch oder tief sein, was dazu führt, daß die Emitter-Kollektor-Strecke leitend oder nichtleitend ist, wie wir in Kapitel 3 gesehen

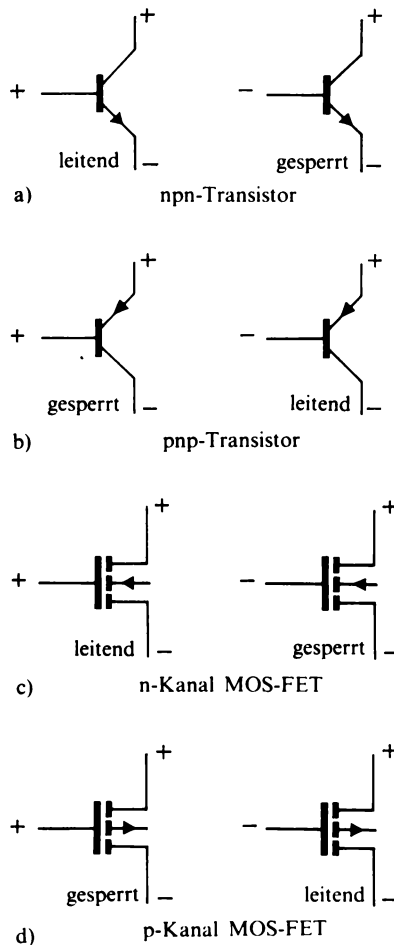


Abb. 22 Schaltsymbole für bipolare Transistoren (a, b) und MOS-FETs vom Anreicherungstyp (c, d)

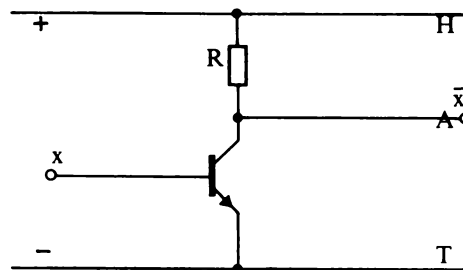


Abb. 23 Negationsschaltung mit bipolarem Transistor

haben. Die Schaltsymbole und -situationen für den MOS-FET vom Anreicherungstyp sind im unteren Teil der Abb. 22 dargestellt. Bei der Negationsschaltung in Abb. 23 ist der Eingangsimpuls die Basis-Emitter-Spannung x. Liegt eine positive Spannung an der Basis an, ist der Transistor leitend. Es fließt ein Strom über den Widerstand R und den Transistor, z ist nur eine sehr geringe Spannung (gegenüber T), d. h. $z = 0$. Liegt keine positive Spannung bei x an, so fließt kein Strom durch R, und z ist die volle Spannung, d. h. $z = 1$. Die Spannung z ist also die Negation der Spannung x: $z = \bar{x}$. Es ist nun überhaupt kein Problem mehr, auch die NOR- und NAND-Schaltung mit rein elektronischen Bauelementen zu realisieren. Man hat dazu im wesentlichen nur die mechanischen Schalter in Abb. 21 durch Transistoren zu ersetzen. Das ist in Abb. 24 geschehen. Wird außerdem der Widerstand R statt hinter den Transistoren (in der Richtung des Stromflusses) vor den Transistoren angebracht, wie das in Abb. 24 der Fall ist, so wird zusätzlich zur OR- bzw. AND-Verknüpfung noch eine Negation ausgeführt. Man erhält also

Logik-Gatter

Schaltungen für die Operationen NOR und NAND. Die Schaltungen in Abb. 23 und Abb. 24 für die elementaren logischen Operationen nennt man auch *Gatter*. Man spricht von NOR- und NAND-Gattern. In Abb. 23 und Abb. 24 sind die *einfachsten* Schaltungen dargestellt, die sich zur Realisierung der elementaren logischen Operationen denken lassen. Der Vergleich der Gatterschaltungen mit den Schaltsymbolen

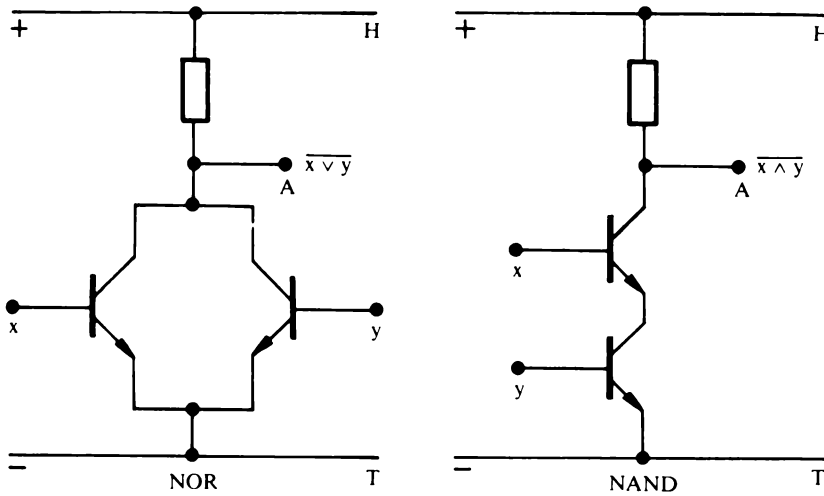


Abb. 24 Schaltungen für die NOR- (a) und NAND-Operation (b) mit Hilfe von bipolaren Transistoren

für die logischen Operationen in Abb. 20 zeigt, daß bei den Gattern zwei zusätzliche Anschlüsse vorhanden sind, nämlich die für die Versorgungsspannung. Sie führen der Schaltung den Informationsträger, die elektrische Spannung, zu, auf den durch die Eingänge Informationen aufgeprägt werden und von dem die verarbeiteten Informationen am Ausgang wieder abgenommen werden. Die Versorgungsspannung hat einen Versorgungsstrom zur Folge. Dieser fließt bei den Schaltungen in Abb. 23 und 24 in allen Transistoren im Emitter-Kollektor-Kreis. Er beläuft sich auf einige mA und ist damit relativ stark, verglichen jedenfalls mit dem Schaltstrom, der im Emitter-Basis-Kreis fließt und nur einige Mikroampere beträgt. Außer zum Schalten verbrauchen die Logik-Gatter in Abb. 23 und 24 also auch im Ruhestand, ohne daß geschaltet wird, elektrische Leistung. Da die Spannung zwischen Emitter und Kollektor von der Größenord-

nung einiger Volt ist, beträgt sie einige Milliwatt pro Gatter und ist damit wesentlich größer als die Schaltleistung eines Transistors, für die wir im Kapitel 3 einige Mikrowatt abgeschätzt hatten. Auch die Verzögerungszeit, die zwischen der Eingabe der Spannungsimpulse in die Eingänge eines Gatters und dem Auftreten der verarbeiteten Impulse am Ausgang besteht, liegt oberhalb der Verzögerungszeit beim einzelnen Transistor. Sie ist u. a. davon abhängig, wie schnell die Spannung zwischen den einzelnen Transistoren innerhalb der Schaltung übertragen werden kann. Die dafür erforderliche Zeit ist durch den Widerstand und die Kapazität der Bauelemente und der Verbindungsleitungen bestimmt. In mikroelektronischen Schaltungen müssen die Logik-Gatter miteinander gekoppelt werden, d. h., die Ausgänge der einen Gatter müssen an die Eingänge der anderen Gatter angelegt werden. Dazu ist es erforderlich, die Eingangs- und Ausgangsspannungen

bzw. -widerstände der Gatter aneinander anzupassen. Das macht in der Regel zusätzliche Widerstände und Transistoren erforderlich. Schnelle Gatter müssen möglichst kleine Widerstände und kleine Kapazitäten haben. Mit der Verkleinerung des Widerstandes nimmt aber automatisch der Leistungsverbrauch zu. Das heißt, daß sich Verzögerungszeit und Leistungsverbrauch eines Gatters gegenläufig verhalten. Die Schnelligkeit eines Gatters kann dadurch erhöht werden, daß es mit einer größeren Leistung betrieben wird, indem eine höhere Spannung angelegt wird. Verzögerungszeit und Leistungsverbrauch eines Gatters sind also innerhalb gewisser Grenzen variabel. Die Güte G eines Gatters wird gewöhnlich durch das Produkt aus Leistungsverbrauch und Verzögerungszeit charakterisiert. Für die Gatter in Abb. 23 und 24 können wir einen Leistungsverbrauch von 10 mW annehmen, so daß sich für G der Wert 1000 pJ ($1 \text{ pJ} = 10^{-12} \text{ Joule}$) ergibt. Der Vergleich mit der Schaltenergie eines einzelnen Transistors, die wesentlich unter 1 pJ liegt, gibt einen Hinweis darauf, wie schlecht wir die guten Eigenschaften des einzelnen Transistors in den Gattern aus Abb. 23 ausnutzen. Diese Gatter sind zwar einfach, aber sie verbrauchen zuviel Leistung und sind zu langsam. Man muß versuchen, schnellere Gatter zu entwickeln, die weniger Leistung verbrauchen und dabei gleichzeitig auch noch andere positive Eigenschaften besitzen, wie z. B. gute Realisierbarkeit und hohe Betriebssicherheit. In der Entwicklung solcher Gatter besteht ein wichtiges Ziel der Mikroelektronik. Bei seiner

Verfolgung ist eine Reihe von Schaltungstypen für die elementaren logischen Operationen entwickelt worden.

Schaltungsfamilien

Man bezeichnet sie als *Schaltungslogiken* oder *Schaltungsfamilien*. Die Schaltungen in Abb. 24 nennt man «Direkt gekoppelte Transistor-Logik» (englisch: Direct Coupled Transistor Logic) und kürzt sie als DCTL ab. Die DCTL-Schaltungen kommen aus den oben erwähnten Gründen heute nicht mehr zum Einsatz. Die heute verwendeten Logiken auf der Basis von bipolaren Transistoren sind vor allem die TTL (Transistor-Transistor-Logic), die I²L (Integrated Injection Logic) und die ECL (Emitter Coupled Logic). Das Grundgatter der TTL-Schaltungsfamilie, ein NAND-Gatter, ist in Abb. 25 dargestellt. Im Unterschied zu den Gattern in Abb. 23 und 24, wo die Eingangsinformation auf die Basiselektroden der beiden Transistoren aufgegeben wird, liegen beim TTL-Gatter die Eingangsspannungen an den beiden Emittenelektroden eines 2-Emitter-Transistors T₁. Befindet sich wenigstens einer der beiden Eingänge auf niedrigem Potential, ist also $x = 0$, $y = 0$ oder $x = 1$, $y = 0$ oder $x = 0$, $y = 1$, so wird die Emitter-Basis-

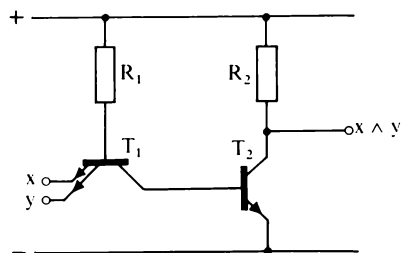


Abb. 25 NAND-Gatter in Transistor-Transistor-Logik (TTL)

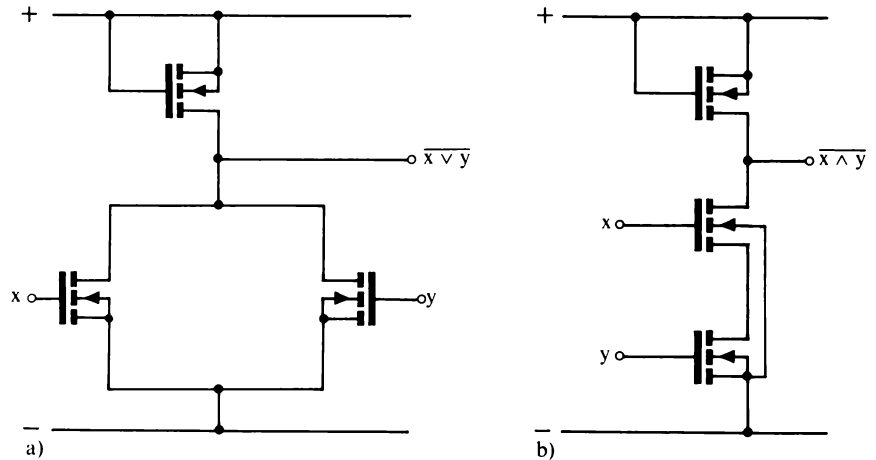


Abb. 26 NOR- (a) und NAND-Gatter (b) in n-MOS-Logik

Strecke von T₁ leitend. Durch den Widerstand R₁ und die Emitter-Basis-Strecke von T₁ fließt ein Strom. Basis und Kollektor von T₁ befinden sich auf niedrigem Potential. Dadurch wird der Transistor T₂ gesperrt, über R₂ fällt praktisch keine Spannung ab. Das Potential hinter R₂, das auf den Ausgang übertragen wird, ist hoch. Bei $x = 0$, $y = 0$ und $x = 0$, $y = 1$ und $x = 1$, $y = 0$ entsteht am Ausgang also 1, d. h. $\overline{x \wedge y}$. Ist $x = 1$, $y = 1$, so wird die Emitter-Basis-Strecke von T₁ gesperrt. Das Potential der Basis und damit auch des Kollektors von T₁ ist hoch. Der Transistor T₂ wird leitend. Das Potential hinter R₂ wird niedrig, am Ausgang entsteht 0, wie es für die NAND-Verknüpfung erforderlich ist. Neben zwei Emittenelektroden werden in TTL-Gattern auch drei und mehr eingesetzt. Verwendet man bei der Realisierung der Gatter anstelle von bipolaren Transistoren MOS-Feldeffekttransistoren, so spricht man von MOS-Logiken. Ein Beispiel dafür ist in

Abb. 26 dargestellt. Die Schaltungen in Abb. 26 werden dadurch erhalten, daß die DCTL-Schaltungen aus Abb. 24 mit Hilfe von n-Kanal-MOS-FETs vom Anreicherungstyp realisiert werden und außerdem als Widerstände ebenfalls n-Kanal-MOS-FETs benutzt werden. Den verschiedenen Arten von MOS-FETs entsprechen verschiedene MOS-Schaltungsfamilien. Werden n- und p-MOS-FETs vom Anreicherungstyp in ein und demselben Schaltkreis verwendet, so spricht man von C (Complementary) MOS-Schaltungen. Eine detaillierte Behandlung der verschiedenen Schaltungsfamilien würde über den Rahmen einer ersten Einführung in die Mikroelektronik hinausgehen. Es kann darauf um so eher verzichtet werden, als es dabei nicht um grundsätzlich Neues geht. Das Wesentliche zur Realisierung der elementaren logischen Operationen mittels elektronischer Schaltungen haben wir uns bereits an den behandelten Beispielen klarmachen können.

Über die Güte der heute am häufigsten verwendeten Schaltungsfamilien gibt die Abb. 27 Auskunft. Dort sind die von einem Gatter verbrauchte elektrische Leistung und die Verzögerungszeit, d. h. die für eine logische Operation erforderliche Zeit, dargestellt. Die verschiedenen Schaltungsfamilien sind als umrandete Gebiete eingezeichnet, deren Ordinatenwerte dem Leistungsverbrauch und deren Abszissenwerte den Verzögerungszeiten entsprechen. Wir entnehmen aus Abb. 27 z. B., daß bei der TTL-Schaltung Verzögerungszeiten unter 10 ns bei einem Leistungsverbrauch zwischen 1 und 10 mW erreicht werden. Etwa den gleichen Leistungsverbrauch, aber Verzögerungszeiten von einigen 10 bzw. einigen 100 ns weisen die n-MOS- und die p-MOS-Schal-

tungen auf. Bei annähernd gleichen Verzögerungszeiten ermöglicht die CMOS-Logik einen etwa 100fach geringeren Leistungsverbrauch, die I²L-Schaltung sogar einen 1000fach geringeren. Daß in Abb. 27 nicht jeder Logik genau ein Punkt zugeordnet wird, sondern ein Gebiet bestimmter Größe, hat verschiedene Gründe. Einen haben wir bereits kennengelernt — Leistungsverbrauch und Verzögerungszeit sind durch die Wahl der Betriebsspannung in gewissen Grenzen variabel. Ein anderer Grund sind die verschiedenen Realisierungstechniken, die bei der Herstellung der Gatter angewandt werden. Ausführlicher werden wir auf die Herstellungstechnologie in Kapitel 5 eingehen. Hier bemerken wir nur soviel, daß Schaltungslogik und Herstellungstechno-

logie eng miteinander zusammenhängen. Bestimmte Logiken bzw. Schaltungsfamilien sind zum Teil an ganz bestimmte Herstellungstechniken gebunden (zum Beispiel MOS-Logiken an MOS-Technik). Mitunter wird die Unterscheidung zwischen Schaltungslogik und Realisierungstechnik überhaupt ignoriert und von Technik gesprochen, wenn eigentlich die Schaltungslogik gemeint ist.

Speicherschaltungen

Neben den elementaren logischen Operationen mit 0- und 1-Impulsen ist es vor allem das *Speichern* von Information in Form solcher Impulse, wofür Schaltungen benötigt werden. Soll sich eine Schaltung für diesen Zweck eignen, so muß sie zwei stabile Zustände besitzen, die über einen Eingang von außen wahlweise eingestellt werden können. Außerdem muß durch den Spannungswert am Ausgang feststellbar sein, in welchem Zustand sich die Schaltung befindet. Dieser Spannungswert darf also nicht, wie bei den Logik-Gattern, von der Spannung abhängen, die zur gleichen Zeit am Eingang anliegt. Er muß durch den *Zustand* der Schaltung bestimmt sein, also durch die Spannung, die zur Einstellung dieses Zustandes *in der Vergangenheit* am Eingang *angelegen* hat. Speicherschaltungen unterscheiden sich damit wesentlich von Schaltungen für die elementaren logischen Operationen und anderen Schaltungen, bei denen die Spannung am Ausgang eindeutig durch die gleichzeitig anliegende Spannung am Eingang bestimmt ist. Letztere heißen *kombinatorische* Schaltungen. Zum Speichern sind kombinatorische Schaltungen nicht geeignet, es sei denn, man begreift die spezielle

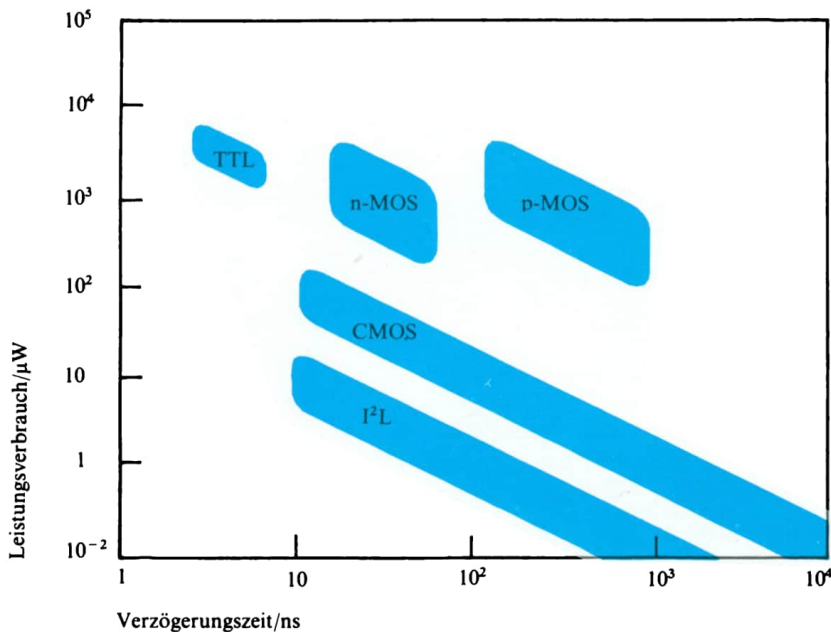


Abb. 27 Leistungsverbrauch und Verzögerungszeit von Logik-Gattern aus verschiedenen Schaltungsfamilien

Art der Zusammenschaltung der Bauelemente bei der Herstellung eines Schaltkreises auch als «Speichern». In diesem Sinne können dann auch kombinatorische Schaltungen Speicher sein, aber nur sogenannte Festwertspeicher, weil eine Änderung des Speicherinhalts bei ihnen nicht möglich ist. Bei den eigentlichen Speicherschaltungen kann die gespeicherte Information geändert werden, ohne daß die Schaltung selbst verändert wird. Geändert wird nur ihr Zustand. Die Einstellung des Zustandes bedeutet das Einschreiben der Information, die Feststellung des Zustandes das Ablesen. Ohne äußere Einwirkung muß die Schaltung möglichst lange in ihrem beim Einschreiben eingestellten Zustand verharren. Dieses Verharren bedeutet das Speichern. In der Mikroelektronik werden für das Speichern zwei verschiedene Grundschaltungen verwendet, die beide die geforderten Eigenschaften besitzen. Dabei handelt es sich zum einen um die sogenannte Flipflop-Schaltung und zum anderen um eine Schaltung mit einem Kondensator als wesentlichem Bauelement.

Flipflop-Schaltung

Wir erläutern zunächst die Flipflop-Schaltung. Sie ist in Abb. 28 schematisch dargestellt. Zum besseren Verständnis denken wir uns zunächst die beiden Transistoren links und rechts außen weg. Bei den verbleibenden beiden inneren Transistoren ist wesentlich, daß die Basis des einen Transistors mit dem Kollektor des anderen verbunden ist und umgekehrt. Dadurch wird erreicht, daß jeweils nur einer der beiden Transistoren leitend sein kann: Ist der linke leitend, so ist der rechte ge-

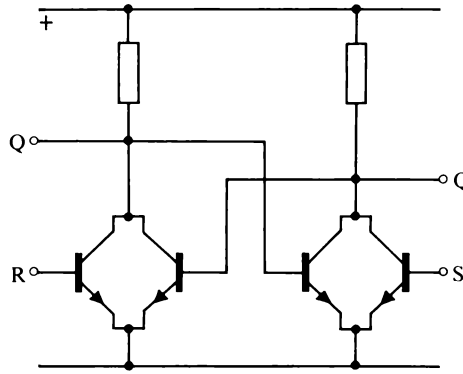


Abb. 28 Flipflop-Schaltung

sperrt. Denn bei leitendem linken Transistor fällt am linken Widerstand die gesamte Spannung zwischen dem hohen «+»-Potential und dem niedrigen «—»-Potential ab, d. h., die Basis des rechten Transistors befindet sich auf niedrigem Potential. Als npn-Transistor ist er dann aber gesperrt (siehe Abb. 22). Ist dagegen der rechte Transistor leitend, so ist der linke gesperrt. Befindet sich die Schaltung einmal in einem der beiden Zustände, so verbleibt sie auch in ihm. Jeder der beiden Zustände ist stabil. Welcher der beiden Zustände sich einstellt, ist durch die Schaltung in Abb. 28 nicht festgelegt. Es sind beide möglich. Die Aussage, daß ein bestimmter Zustand — etwa «links leitend, rechts gesperrt» — vorliegt, ist eine Information mit einem Gehalt von 1 bit. Mit Hilfe der Flipflop-Schaltung kann also 1 bit Information gespeichert werden. Zum Einschreiben der Information dienen die beiden Transistoren links und rechts außen in Abb. 28, die wir nun in die Betrachtung mit einbeziehen. Wird an der Basis des linken Transistors (Eingang R) eine kurze Zeit lang

eine positive Spannung 1 angelegt, so sind danach die beiden linken Transistoren leitend und die beiden rechten gesperrt. Sofern die Schaltung schon vorher in diesem Zustand war, bleibt sie in ihm, war sie im anderen Zustand (links gesperrt, rechts leitend), so ändert sie ihn. In jedem Falle ist sie nach dem Anlegen des 1-Impulses bei R im Zustand «links leitend, rechts gesperrt». Der Wert 1 ist in die Flipflop-Schaltung eingeschrieben worden. Der Zustand der Schaltung kann an den beiden Ausgängen Q und $\bar{Q}$ abgelesen werden. Je nachdem, ob die linke oder rechte Seite leitend ist, befindet sich $\bar{Q}$ auf einem hohen Potential (1) oder auf einem niedrigen (0). Indem man dieses Potential abliest, kann man feststellen, ob 1 oder 0 gespeichert ist.

Kondensatorschaltung

Genau so gut wie mit Hilfe einer Flipflop-Schaltung lassen sich die beiden erforderlichen Zustände mit Hilfe eines Kondensators realisieren. In Abb. 29 ist die entsprechende Speicherschaltung dargestellt. Sie

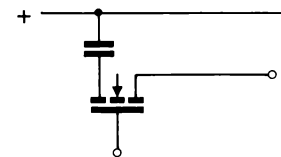


Abb. 29 Speicherschaltung mit Kondensator

enthält außer dem Kondensator noch einen Transistor, der aber mit dem Speichern nichts zu tun hat. Geladener Kondensator bedeutet 1, ungeladener 0. Ist der Kondensator geladen und der Transistor leitend, kann die Ladung über den Aus-

gang abfließen, d. h., eine gespeicherte 1 wird ausgegeben. Ist er nicht geladen, kann nichts abfließen, d. h., eine gespeicherte 0 wird ausgegeben. Über den Ausgang kann der Ladungszustand des Kondensators aber auch geändert werden, d. h., es kann Information eingeschrieben werden. Auf die Funktion des Transistors gehen wir weiter unten noch genauer ein. So einfach wie in Abb. 29 ist die Speicherung mit Hilfe eines Kondensators allerdings doch nicht. Das Problem besteht darin, daß sich der Kondensator durch die unvermeidlichen Verlustströme mehr oder weniger schnell von selbst entlädt und daß beim Ablesen des Ladungszustandes etwas von der Ladung des Kondensators über die Leseleitung abgeführt werden muß. Die gespeicherte Ladung muß deshalb von Zeit zu Zeit aufgefrischt werden, wenn die Information nicht verlorengehen soll. Dazu sind zusätzliche Bauelemente notwendig, die die einfache Kondensatorschaltung wieder komplizierter machen. Als Kondensator wird meist ein MOS-FET verwendet.

Schaltungen zum Speichern von 1 bit bezeichnet man als *Speicherzellen*. Die Schaltung in Abb. 28 ist ein Beispiel für eine *statische Zelle*, die in Abb. 29 für eine *dynamische Zelle*. Die Bezeichnung «dynamisch» deutet darauf hin, daß der eingeschriebene Zustand zeitlichen Änderungen unterliegt, wie wir das oben erläutert haben. Aus den Speicherzellen lassen sich Schaltungen mit größerer Speicherkapazität aufbauen. Dabei ist man bestrebt, mit möglichst wenig äußeren Anschlüssen auszukommen. Würde man zu jeder Speicherzelle eine eigene

Schreib- und Leseleitung hinführen, so würde man bei einem 16bit-Speicher $2 \cdot 16 = 32$ Anschlüsse benötigen. Es gibt aber eine andere Lösung, bei der man mit wesentlich weniger Anschlüssen auskommt. Man verwendet nur eine einzige Schreib- und eine einzige Leseleitung. Mit diesen Leitungen sind alle Zellen verbunden. Da aber immer nur bei einer einzigen Zelle eingeschrieben bzw. ausgelesen werden soll, und nicht bei allen gleichzeitig, muß in der Speicherschaltung die Möglichkeit vorhanden sein, eine bestimmte Zelle auszuwählen oder, wie man auch sagt, zu *adressieren*.

Adressierung von Speicherzellen
Dazu ordnet man die Speicherzellen in Form einer quadratischen Matrix an. Jede Zelle befindet sich im Kreuzungspunkt einer bestimmten Spalte und einer bestimmten Zeile der Matrix. Um die gewünschte Zelle zu finden, muß man die Nummer ihrer Spalte und die Nummer ihrer Zeile kennen. Bei $N \cdot N$ Speicherzellen benötigt man $N + N$ Anschlüsse, um eine bestimmte Zelle zu adressieren. Bei 16 Zellen sind das 8 Anschlüsse, mit dem Schreib- und Leseanschluß also 10, d. h. wesentlich weniger als bei einer eigenen Schreib- und Leseleitung für jede Zelle. Die Zahl der Anschlüsse für die Adressierung kann dadurch noch weiter reduziert werden, daß die Zeilen- und Spaltennummern codiert werden. Im Binärcode benötigt man bei einer 4×4 Speichermatrix statt 2×4 nur $2 \times \log_2 4$ Anschlüsse. Von der Codierung der Adressen wird heute bei allen Speicherschaltungen Gebrauch gemacht. Wie läßt sich nun die Adressierung durch eine elektronische Schaltung realisieren? Wir

wollen annehmen, daß eine Speicherzelle durch einen 1-Impuls an einem zusätzlich an ihr angebrachten Eingang für das Einschreiben oder Auslesen bereit gemacht wird, während ein 0-Impuls an diesem Eingang die Zelle für das Einschreiben oder Auslesen sperrt. Das läßt sich z. B. dadurch erreichen, daß der Schreibeingang bzw. der Leseausgang der Speicherzelle über einen MOS-Transistor mit der Schreib- bzw. Leseleitung verbunden wird, so wie das in Abb. 29 bereits geschehen ist. Die Gate-Elektrode dieses Transistors ist dann der zusätzliche Eingang. Wird der Transistor durch die Gate-Spannung in den leitenden Zustand geschaltet, kann eingeschrieben oder ausgelesen werden. Ist der Transistor gesperrt, kann kein Einschreiben oder Auslesen stattfinden. Die Schaltung zur Adressierung der solcherart erweiterten Zellen kann mit Hilfe von AND-Schaltungen aufgebaut werden. Zu jeder Speicherzelle gehört ein AND-Glied. Sein Ausgang ist mit der Gate-Elektrode des zusätzlichen Transistors der Zelle verbunden. In Abb. 30 ist eine (der Ein-

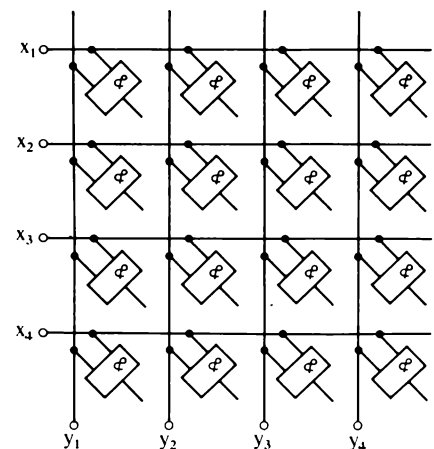


Abb. 30 Adressierung von Speicherzellen

fachheit halber nicht binär codierte) Adresserschaltung für einen 16bit-Speicher mit Hilfe von Logik-Schalt-symbolen dargestellt. Soll die Speicherzelle in Spalte 3, Reihe 2 angesprochen werden, so muß man auf die Anschlüsse x_2 und y_3 jeweils einen 1-Impuls geben und auf alle anderen Anschlüsse den Impuls 0. Dann entsteht am Ausgang eines einzigen AND-Gatters in Abb. 29 ein 1-Impuls, nämlich genau in dem an Spalte 3, Reihe 2. Bei der zugehörigen Speicherzelle ist das Ein- bzw. Auslesen möglich, bei allen anderen nicht. Neben dem soeben beschriebenen Speicher mit *wahlfreiem* Zugriff zu jeder Speicherzelle gibt es auch Speicher mit *seriellem* Zugriff. Man stelle sich dazu vor, daß die Speicherzellen in einer Zeile angeordnet sind. Verbindung zur Schreib-Lese-Leitung hat nur die Zelle am äußersten rechten Ende. Sorgt man dafür, daß der Inhalt der einzelnen Speicherzellen in einem bestimmten zeitlichen Rhythmus in die jeweils benachbarte Zelle geschoben wird — die Zelle ganz rechts gibt ihren Inhalt an die Zelle links außen ab — so liegen in einem vollen Durchlaufzyklus der Reihe nach alle Speicherinhalte genau einmal an der Schreib-Lese-Leitung an und können ausgelesen oder verändert werden. Besondere Bedeutung für die Realisierung serieller Speicher besitzen MOS-Strukturen. Wie bereits erwähnt, kann Information in Form von Ladung in einem Kondensator gespeichert werden. Anstelle eines Kondensators kann auch ein MOS-FET verwendet werden. Die Ladung wird dabei unter der Gate-Elektrode gespeichert. Um einen seriellen Speicher zu realisieren, benötigt man ein

MOS-FET mit vielen in einer Zeile angeordneten Gate-Elektroden. Durch Spannungsimpulse an diesen Elektroden kann die gespeicherte Ladung von einem Gate zum anderen transportiert werden. Man spricht dann von ladungsgekoppelten Bauelementen oder von CCDs (englisch: Charge Coupled Devices).

Zählschaltungen

Zu den Grundoperationen, die an den 0- und 1-Impulsen bei der Informationsverarbeitung ausgeführt werden müssen, gehört auch das Zählen. Man kann das Zählen von Impulsen in einer Impulsfolge als eine spezielle Form des Speicherns auffassen. Zählschaltungen sind daher den Speicherschaltungen sehr ähnlich. Das Zählen bis 2 ist mit einer Schaltung möglich, die in Abb. 31 dargestellt ist. Wir haben dabei die logischen Schaltsymbole aus Abb. 20 verwendet. Dadurch wird die Schaltung wesentlich übersichtlicher, obwohl sie auf den ersten Blick noch kompliziert genug erscheint. Die Verwendung logischer

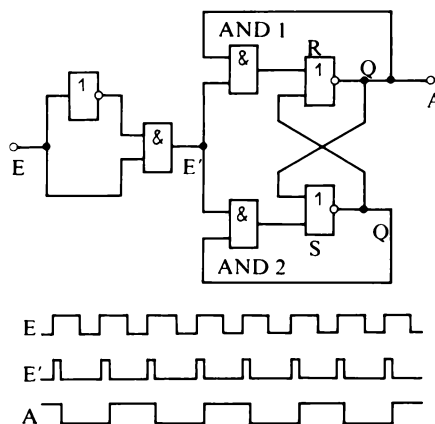


Abb. 31 Zählschaltung. Die Schaltung zählt bis 2.

Schaltungen anstelle der elektrischen ist in der Elektronik ein häufig benutztes Verfahren, um kompliziertere Schaltungen übersichtlich darzustellen.

Bei der Erläuterung der Schaltung in Abb. 31 gehen wir schrittweise vor. Zunächst sehen wir uns noch einmal die einfache Flipflop-Schaltung an, die ein Bestandteil des Schaltkreises in Abb. 31 ist. In Abb. 28 wurde die Flipflop-Schaltung unter Verwendung von elektrischen Schaltsymbolen dargestellt. Werden logische Symbole benutzt, ergibt sich für ein Flipflop die Schaltung in Abb. 32a. Man kann sich sofort klarmachen, daß nur zwei verschiedene Belegungen der Anschlüsse R, S, Q, $\bar{Q}$ mit 0- und 1-Werten logisch möglich sind. Sie sind in Tabelle 3 dargestellt. Die durchge-

Tab. 3 Anschlußbelegungen eines Flipflop. Die durchgestrichenen Werte sind ausgeschlossen.

R	1	0	1	0
S	0	1	0	1
Q	0	1	1	0
$\bar{Q}$	1	0	0	1

strichenen Belegungen führen zu logischen Widersprüchen. Den beiden logisch widerspruchsfreien Belegungen entsprechen die beiden stabilen Zustände der Flipflop-Schaltung. Eine weitere Teilschaltung der Zählschaltung in Abb. 31, deren Funktion wir uns zunächst gesondert klarmachen wollen, ist in Abb. 32b dargestellt. Die Schaltung bildet den logischen Ausdruck $E' = E \wedge \bar{E}$, und dieser ist stets gleich 0. Nach den Regeln der Logik

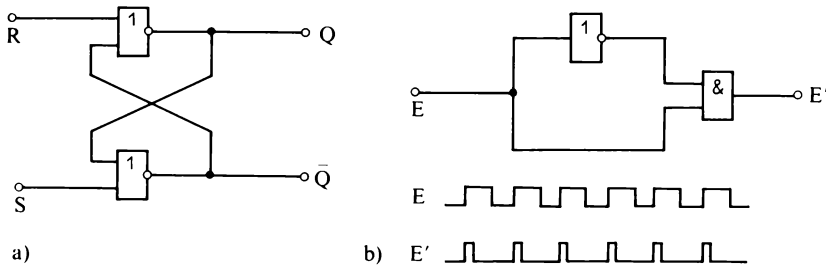


Abb. 32 Logische Schaltung für ein Flipflop (a) und zur Erzeugung kurzer Impulse (b)

entsteht am Ausgang E' also stets 0, gleichgültig, ob der Eingang E mit 0 oder 1 belegt ist. Bei der realen Schaltung werden die logischen Operationen mit Hilfe von Transistoren ausgeführt, und diese arbeiten, wie wir wissen, nicht unendlich schnell, sondern besitzen eine bestimmte Verzögerungszeit. In unserem Falle bedeutet das konkret, daß der negierte Eingangsimpuls etwas später in den oberen Eingang des AND-Gliedes einläuft als der Eingangsimpuls selbst in den unteren. War der obere Eingang vor der Eingabe des Impulses bei E mit 1 belegt und ist $E = 1$, so erhalten wir am Ausgang solange 1, bis der im Negator gebildete 0-Impuls am oberen Eingang des AND-Gliedes anliegt. Von da an ist $E' = 0$. Ist $E = 0$, so ergibt sich am Ausgang immer, d. h. auch während der Verzögerungszeit des Negators, der Wert 0. In Abb. 32b ist die Impulsfolge eingezeichnet, die bei Eingabe einer gleichmäßigen Folge von 0- und 1-Impulsen bei E am Ausgang A in der realen elektronischen Schaltung entsteht. Es ist eine Folge sehr kurzer 1-Impulse, die immer dann auftreten, wenn am Eingang ein 0-Impuls in einen 1-Impuls übergeht, d. h. an den positiven Flanken der Impulsfolge am Eingang.

Damit sind wir jetzt hinreichend vorbereitet, um die Zählerschaltung in Abb. 31 zu verstehen. Wir geben auf den Eingang E eine gleichmäßige Folge von 0- und 1-Impulsen, so wie in Abb. 31 angedeutet. Das Flipflop-Glied in Abb. 31 sei im Zustand mit $R = 0$, $S = 1$, $Q = 1$, $\bar{Q} = 0$. Was wird geschehen? Bei E' tritt die in Abb. 32b gezeigte Folge kurzer 1-Impulse auf. Diese gelangen auf je einen der Eingänge der beiden AND-Glieder. Bei «AND 1» sind dann beim Auftreffen eines solchen kurzen 1-Impulses auf den unteren Eingang beide Eingänge kurzzeitig mit 1 belegt, da am oberen Eingang Q anliegt und $Q = 1$ ist. Am Ausgang von «AND 1» tritt damit kurzzeitig ein 1-Impuls auf. Der kurze 1-Impuls trifft auf den Eingang R des Flipflop, der bisher mit 0 belegt war, und schaltet das Flipflop in den anderen Zustand. Nun ist $R = 1$, $S = 0$, $Q = 0$, $\bar{Q} = 1$. Am Zustand der Schaltung ändert sich so lange nichts, bis bei E' erneut ein kurzer positiver Impuls entsteht. Nun folgt die gleiche Überlegung wie beim ersten kurzen Impuls, nur daß jetzt nicht Q, sondern $\bar{Q} = 1$ ist, was dazu führt, daß nicht am Ausgang von «AND 1», sondern am Ausgang von «AND 2» ein kurzer 1-Impuls auftritt. Dieser 1-Impuls

gelangt auf den S-Eingang der Flipflop-Schaltung, der vorher mit 0 belegt war, und schaltet das Flipflop in den anderen Zustand. Jetzt ist wieder $R = 0$, $S = 1$, $Q = 1$, $\bar{Q} = 0$. Die Ausgangssituation ist wieder hergestellt. Das Spiel kann von neuem beginnen. Dazu waren zwei positive Impulse nötig. Man ahnt bereits, daß am Ausgang A ein positiver Impuls auf je zwei positive Impulse am Eingang entstehen wird. Das ist im unteren Teil von Abb. 31 noch einmal verdeutlicht. Dort ist die Impulsfolge am Eingang E, an E' und am Ausgang A dargestellt. Man erkennt: die Anzahl der Impulse am Ausgang der Schaltung ist halb so groß wie die am Eingang. Nach jedem zweiten Impuls am Eingang entsteht ein Impuls am Ausgang. Die Schaltung zählt bis 2. Durch das Hintereinanderschalten solcher Schaltungen, d. h. durch die Verbindung des Ausganges der vorhergehenden mit dem Eingang der nachfolgenden Schaltung, läßt sich die Impulszahl noch weiter untersetzen. Mit n Schaltungen erreicht man eine Untersezung im Verhältnis von $1:2^n$, d. h., auf 2^n Impulse am Eingang der ersten Schaltung kommt ein Impuls am Ausgang der n-ten (siehe Abb. 33). Die Gesamtschaltung zählt bis 2^n . Mit zehn Schaltungen kann man bis $2^{10} = 1024$ zählen. Es ist klar, daß man auf diese Weise die Schaltung für eine Quarzuhr aufbauen kann. Wir gehen darauf in

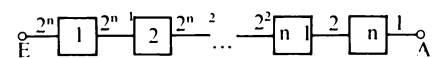


Abb. 33 Zählerschaltung zum Zählen bis 2^n . Die Kästchen symbolisieren die Schaltung aus Abb. 31. Auf 2^n Impulse am Eingang E kommt ein Impuls am Ausgang A.

Kapitel 6 noch etwas genauer ein.

Speicher- und Zählschaltungen sind Beispiele dafür, wie mit Hilfe der elementaren logischen Operationen kompliziertere Funktionen realisiert werden können. In beiden Fällen handelt es sich aber nicht um kombinatorische Schaltungen. Als kombinatorische Schaltungen haben wir bisher nur die für die elementaren logischen Operationen kennengelernt. Im folgenden wollen wir auf einige komplexere kombinatorische Schaltungen eingehen, die für die Informationsverarbeitung praktische Bedeutung besitzen.

Im Zusammenhang mit der Adressierung von Speicherzellen hatten wir **Codier- und Decodierschaltungen** die Codierung von Zahlen erwähnt. Schaltungen, die eine Zahl aus einer Art der Codierung in eine andere, gewünschte Art der Codierung überführen, nennt man *Codierschaltungen*. Leisten sie das Umgekehrte, heißen sie *Decodierschaltungen*. In Abb. 34 ist eine Schaltung für die Decodierung der binär codierten Zeilennummer einer 4×4 Speicher-matrix dargestellt. Aus Tabelle 4 kann man entnehmen, welche Spannungswerte an den beiden Eingängen zu welchen Spannungswerten an den

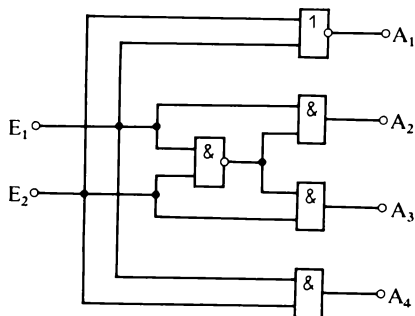


Abb. 34 Decodierschaltung

vier Ausgängen führen. Wir überlassen die Verifikation dem Leser. Bei jeder binär codierten Zahl zwischen 0 und 3 an den Eingängen entsteht an einem der Ausgänge eine 1, an den drei übrigen eine 0. Der Ausgang, an dem die 1 entsteht, ist eindeutig der Binärzahl am Eingang zugeordnet. Durch die Belegung der zwei Eingänge mit einer bestimmten Binärzahl läßt sich genau einer aus den vier Ausgängen auswählen. Man spricht deshalb von einer „1 aus 4“-Darstellung.

Oft entsteht bei der Informationsverarbeitung die Aufgabe, gleichzeitig vorliegende Impulse in mehreren Leitungen in zeitlich aufeinanderfolgende Impulse in einer einzigen Leitung umzuwandeln. Diese **Multiplexer-Schaltung**

gibt es von einer sogenannten *Multiplexer-Schaltung* gelöst. Für das umgekehrte Problem, die Verwandlung einer zeitlichen Aufeinanderfolge von Impulsen in einer Leitung in zeitlich parallele Impulse in mehreren Leitungen ist der *Demultiplexer* zuständig. In Abb. 35 ist ein Multiplexer dargestellt. Die mit T_1, T_2, T_3, T_4 bezeichneten Pole sind für die Taktimpulse bestimmt. Die Form der vier Taktimpulse, zu-

sammen mit der Belegung der Eingänge und der Impulsfolge am Ausgang, ist in Teil b von Abb. 35 dargestellt. Die Taktimpulse bewirken, daß im ersten Takt der bei E_1 anliegende 1-Impuls bis zum Ausgang A durchdringen kann, für die drei anderen Eingänge ist der Weg zum Ausgang gesperrt. Im zweiten Takt öffnet der Taktimpuls von T_2 den Weg für die bei E_2 anliegende 0 zum Ausgang usw. Die zeitlich parallelen Impulse an den vier Eingängen werden also tatsächlich in zeitlich auf-

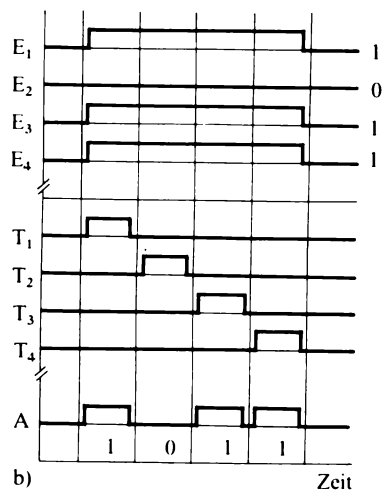
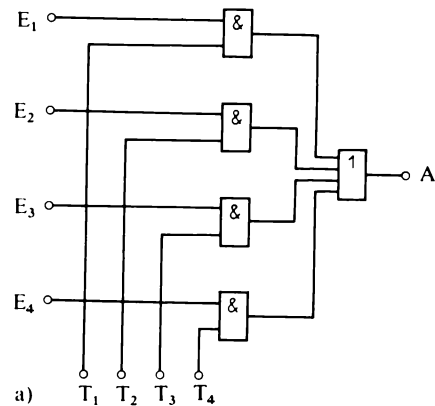


Abb. 35 Multiplexer-Schaltung

Tab. 4 Belegung der Ein- und Ausgänge der Decodierschaltung aus Abb. 34

E_1	E_2	A_1	A_2	A_3	A_4
0	0	1	0	0	0
1	0	0	1	0	0
0	1	0	0	1	0
1	1	0	0	0	1

einanderfolgende an dem einen Ausgang umgewandelt.

Addierschaltungen

Als weiteres praktisch wichtiges Problem behandeln wir schließlich noch die Addition von zwei Binärzahlen. Wir beginnen mit einer Schaltung für die Addition von zwei einstelligen Binärzahlen x, y . Das Ergebnis kann eine zweistellige Binärzahl sein (siehe Tabelle 5).

Tab. 5 Wertetabelle zur Addition zweier einstelliger Binärzahlen

x	y	z 1. Stelle	$+ \ddot{u}$ 2. Stelle
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

Die Ziffer der ersten Stelle bezeichnen wir mit z , die der zweiten mit $\ddot{u}$. Letztere stellt den Übertrag der ersten in die zweite Stelle dar. Die beiden Ziffern erhält man durch folgende Kombination der elementaren logischen Operationen:

$$z = (\bar{x} \wedge y) \vee (x \wedge \bar{y}),$$

$$\ddot{u} = x \wedge y$$

Man überzeugt sich davon durch direktes Verifizieren der Wertetabelle. Ist z.B. $x = 0, y = 1$ so ist zur Berechnung von z gemäß obiger Gleichung folgende Überlegung durchzuführen: $\bar{x} = 1$, also ist $\bar{x} \wedge y = 1, \bar{y} = 0$, also ist $x \wedge \bar{y} = 0, 1 \vee 0 = 1$, das Ergebnis nach obiger Gleichung ist also 1, wie es die Wertetabelle ausweist. Die Berechnung von z gemäß der obigen Gleichung wird durch die logische Schal-

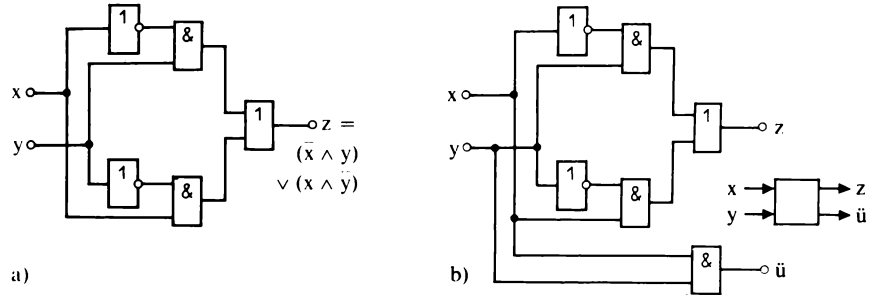


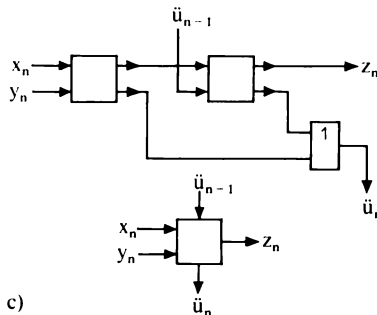
Abb. 36 Addierschaltungen

a) Schaltung zur Berechnung der ersten Stelle der Summe zweier Binärzahlen,

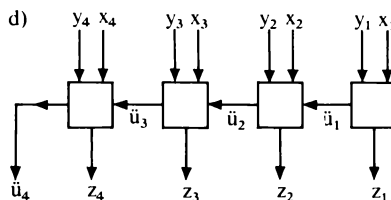
b) Schaltung zur Berechnung beider Stellen (Halbadder),

c) Volladder aus zwei Halbaddern,

d) Schaltung zur Addition von zwei 4stelligen Binärzahlen $x_4x_3x_2x_1$ und $y_4y_3y_2y_1$



c)



d)

tung in Abb. 36 ausgeführt. In Abb. 36 b) haben wir diese Schaltung so erweitert, daß auch noch der Übertrag $\ddot{u}$ berechnet wird. Man bezeichnet diese erweiterte Schaltung als «Halbadder» (HA). Wenn es uns nicht darum geht, die Funktionsweise des Halbadders im einzelnen darzustellen, sondern nur das Ergebnis seiner Wirkung zu veranschaulichen, kann er zu einem Block mit zwei Eingängen und zwei Ausgängen

zusammengefaßt werden, so wie es in Abb. 36 b) zu sehen ist.

Wir wollen jetzt zwei N -stellige Binärzahlen $x_Nx_{N-1} \dots x_1$ und $y_Ny_{N-1} \dots y_1$ bis y_1 addieren. Die Summe bezeichnen wir mit $z_{N+1}z_N \dots z_1$. Es gilt also

$$\begin{array}{r} x_Nx_{N-1} \dots x_1 \\ y_Ny_{N-1} \dots y_1 \\ \hline z_{N+1}z_Nz_{N-1} \dots z_1 \end{array}$$

Bei der Berechnung der Summe ist zu beachten, daß von der zweiten Stelle an nicht nur die beiden Ziffern der jeweiligen Stelle zu addieren sind, sondern auch der Übertrag von der jeweils vorhergehenden Stelle. Betrachten wir z. B. die n -te Stelle, so ist z_n die Ziffer der ersten Stelle der Summe $x_n + y_n + \ddot{u}_{n-1}$. Dabei bezeichnet $\ddot{u}_{n-1}$ den Übertrag von der $(n-1)$ -ten Stelle. Mit Hilfe von zwei Halbaddern läßt sich z_n leicht berechnen. Die entsprechende Schaltung ist in Abb. 36 c) dargestellt. Um den Übertrag $\ddot{u}_n$ zu erhalten, brauchen wir nur den Übertrag am Ausgang des 1. HA zum Übertrag des 2. HA zu addieren, so wie das in Abb. 36 c) geschieht. Diese Schaltung bezeichnet man als Volladder und

vereinfacht sie, wenn wieder nur das Ergebnis der Wirkung interessiert, durch die in Abb. 36c gezeigte Schaltung mit drei Eingängen und zwei Ausgängen.

In Abb. 36d zeigen wir die Addition einer vierstelligen Binärzahl. Der Übertrag $\bar{u}_4$ ist entweder $0 \cdot 2^4$ (also kein Übertrag) oder $1 \cdot 2^4 = 16$. Soll im binär codierten Dezimalsystem gerechnet werden, benötigt man als Übertrag von einer Tetrade zur nächsten aber nicht 16, sondern 10. Hier führt nun die parallele Verwendung des Binär- und Dezimalsystems erstmalig zu Schwierigkeiten. Die beiden Zahlensysteme stoßen hart aufeinander, und es sind einige zusätzliche Operationen nötig, um sie miteinander auszusöhnen. Wir wollen aber darauf im einzelnen nicht eingehen.

Betrachten wir das relativ übersichtliche Schaltbild in Abb. 36d und stellen uns gleichzeitig vor, daß anstelle von Voll- und Halbaddern die detaillierten Schaltbilder bis hin zu den einzelnen Negations-, NOR- und NAND-Schaltungen gezeichnet worden wären, so wird der Vorteil der Schaltblöcke deutlich. Mit Hilfe solcher Schaltblöcke lassen sich Stufe für Stufe komplizierte Schaltungen aufbauen (die Schaltung in Abb. 36d enthält 52 logische Operationen und mehrere 100 Bauelemente), ohne daß man die Kompliziertheit auf den einzelnen Stufen zu spüren bekommt. Auf jeder Stufe treten nur relativ einfache Schaltungen auf. Um die elektronische Realisierung brauchen wir uns nur auf der niedrigsten Stufe, der der elementaren logischen Operationen, zu kümmern, weil sich die höheren Stufen aus Elementen der niedrigeren Stufe aufbauen.

Analogschaltungen

Wir haben uns bisher auf Schaltungen für die digitale Informationsverarbeitung als die der Mikroelektronik besonders gut angepaßte Form beschränkt. Aber auch für die analoge Informationsverarbeitung gibt es heute integrierte Schaltkreise. In einem Buch über Mikroelektronik dürfen deshalb auch Analogschaltungen nicht fehlen. Typische Analogschaltungen, die als mikroelektronische Schaltkreise zur Verfügung stehen, sind Schaltungen für die Verstärkung analoger Signale. Immer größere Bedeutung erlangen auch mikroelektronische Schaltkreise, durch die analoge Signale in digitale, und umgekehrt digitale in analoge umgewandelt werden können. Man bezeichnet sie als AD- bzw. DA-Wandler. Es ist klar, daß AD- und DA-Wandler-Schaltungen neben digitalen immer auch analoge Teilschaltungen enthalten müssen. Auf eine spezielle AD-Wandler-Schaltung wollen wir im folgenden genauer eingehen. Sie ist in Abb. 37 dargestellt.

Analog-Digital-Wandler

Der analoge Teil dieser sonst digitalen Schaltung ist der Komparator, der in Abb. 37a als Dreieck mit zwei Eingängen und einem Ausgang gezeichnet ist. Abb. 37b veranschaulicht seine Wirkungsweise. Der Komparator vergleicht die an den beiden Eingängen E und R anliegenden Spannungen U_E und U_R . Ist U_E größer als U_R , entsteht am Ausgang die positive Spannung U_S , ist U_E kleiner als U_R , die negative Spannung $-U_S$. Wichtig ist, daß die Ausgangsspannung U_A nur dem Vorzeichen nach, nicht aber der absoluten Größe nach, von der Differenz

der Eingangsspannungen abhängt. Die absolute Größe ist immer dieselbe, gleich, wie groß oder von welchem Vorzeichen die Eingangsspannungsdifferenz ist. Der Ausgang des Komparators kann nur zwei Zustände einnehmen, nämlich $+U_S$ und $-U_S$, die als logische 1 für $+U_S$ und logische 0 für $-U_S$ weiter verwendet werden können. Komparatoren sind ihrem Wesen nach Differenzverstärker mit sehr großer Verstärkung, aber begrenzter Ausgangsspannung. Sie lassen sich deshalb auch mit Hilfe von Verstärkern realisieren.

Nachdem die Wirkungsweise des Komparators im Prinzip geklärt ist, läßt sich die Funktion der Gesamtschaltung zur Analog-Digital-Wandlung in Abb. 37a leicht verstehen. Die Schaltung realisiert fast wortgetreu das Verfahren, das wir am Anfang dieses Kapitels zur Gewinnung von digitalen Impulsfolgen aus einer analogen Folge angegeben haben. Wir empfehlen dem Leser, sich dieses Verfahren zunächst noch einmal in Erinnerung zu rufen, weil wir uns jetzt direkt darauf beziehen werden. Die Einteilung der Spannungsskala in einzelne Intervalle erfolgt durch die Hintereinanderschaltung von Widerständen. Die Spannungswerte am Ausgang der vier Widerstände in Abb. 37a sind von oben nach unten $5/6 U_0$, $3/6 U_0$ und $1/6 U_0$. Die Spannung U_0 muß so groß gemacht werden, daß die Spannungswerte U in der analogen Folge alle kleiner sind als U_0 . Andererseits sollte U_0 nicht unnötig groß sein, sondern bei Wahrung der Bedingung $U < U_0$ möglichst klein gemacht werden, damit wir bei der Diskretisierung der Spannungswerte nicht unnötig Ge-

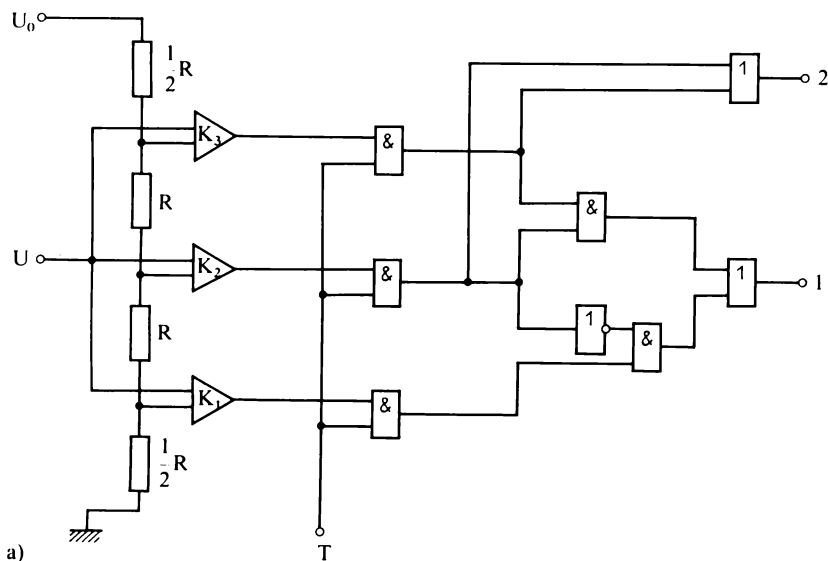
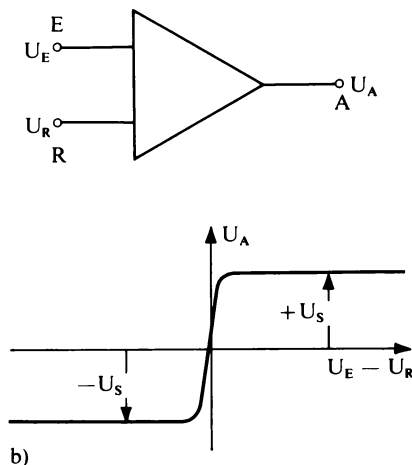


Abb. 37 AD-Wandler-Schaltung

a) Prinzipschaltung für die parallele Analog-Digital-Wandlung. Die Schaltung wandelt das Analogsignal in eine zweistellige Binärzahl um.

b) Zur Funktionsweise eines Komparators



naugigkeit verschenken. Am Komparator 3 in Abb. 37 wird festgestellt, ob U größer als $5/6 U_0$ ist. Wenn das zutrifft, entsteht am Ausgang eine 1, wenn nicht, eine 0. Am Ausgang des Komparators 2 erscheint eine 1, wenn $U > 3/6 U_0$, eine 0, wenn $U < 3/6 U_0$. Der Ausgang des Komparators 1 ist mit einer 1 belegt, wenn $U > 1/6 U_0$ und mit einer 0,

wenn $U < 1/6 U_0$ ist. Auf diese Weise wird die Spannung U in eines der vier Spannungsintervalle in Tabelle 6, erste Spalte, eingeordnet. Die jedem Intervall entsprechende Belegung der Komparatorausgänge ist in der dritten Spalte aufgeschrieben. Rundet man im obersten Spannungsintervall auf, im untersten ab und nimmt in den beiden inneren den

Mittelwert, so entstehen die für die vier Intervalle repräsentativen Spannungswerte in der zweiten Spalte von Tabelle 6. Eine am Eingang anliegende Spannung U wird als einer dieser vier Werte, also als $3 \cdot U_0/3$, $2 \cdot U_0/3$, $1 \cdot U_0/3$ oder als $0 \cdot U_0/3$ identifiziert. Man braucht nun nur noch die Zahlen 3, 2, 1, 0 als zweistellige Binärzahlen zu codieren, und der AD-Wandler ist fertig. Die Codierung geschieht durch die Logik-Schaltung hinter den Komparatorausgängen oder genauer, hinter den AND-Gliedern, die diesen Ausgängen folgen und auf deren Funktion wir gleich noch eingehen werden. Der Leser kann sich leicht selbst davon überzeugen, daß an den beiden Binärausgängen y_2 (zweite Binärstelle) und y_1 (erste Binärstelle) der Logik-Schaltung die richtigen, d. h. die in Spalte 4 von Tabelle 6 aufgeschriebenen Werte, entstehen. Unser AD-Wandler in Abb. 37a wandelt ein Analogsignal in eine zweistellige Binärzahl um. Dafür werden drei Komparatoren benötigt. Wünscht man eine höhere Genauigkeit der AD-Umwandlung, muß man mit

Tab. 6 Erläuterung zur AD-Wandler-Schaltung aus Abb. 37

Spannung U am Analogeingang	Gerundeter Spannungswert	Belegung der Komparator- ausgänge			Belegung der Binärausgänge	
		K_3	K_2	K_1	y_2	y_1
5/6 bis $1 U_0$	$3 \cdot U_0/3$	1	1	1	1	1
3/6 bis 5/6 U_0	$2 \cdot U_0/3$	0	1	1	1	0
1/6 bis 3/6 U_0	$1 \cdot U_0/3$	0	0	1	0	1
0 bis 1/6 U_0	$0 \cdot U_0/3$	0	0	0	0	0

engeren Spannungsintervallen und demzufolge mehr Komparatoren arbeiten. Entsprechend höher muß die Stellenzahl der Binärzahl am Ausgang sein. Bei einer Binärzahl von n Stellen oder, was dasselbe ist, von n Bit, benötigt man $2^n - 1$ Komparatoren. Diese Regel ist natürlich auch in unserem Beispiel erfüllt: $2^2 - 1 = 3$. Die AND-Glieder am Ausgang jedes Komparators des AD-Wandlers in Abb. 37a gewährleisten, daß nur dann die digitalisierten Analogsignale auf die Binärausgänge durchgeschaltet werden, wenn auf den Eingang T ein Taktpuls gegeben wird.

Wichtigstes Element des AD-Wandlers in Abb. 37 ist der Komparator. Er ist ein Beispiel für eine Analog-Schaltung. Auf die Behandlung weiterer Analog-Schaltungen wollen wir verzichten. Das ist möglich, ohne daß dadurch etwas verlorengeht, was für das Verständnis der Mikroelektronik unentbehrlich wäre. Nur darf daraus nicht der Schluß gezogen werden, daß die analogen Schaltungen in der Mikroelektronik neben den digitalen nicht wichtig wären. Zwar lassen sich prinzipiell alle elektronisch überhaupt lösbaren Informationsverarbeitungsprobleme auch in digitaler Form lösen, wenn man von der bei der Digitalisierung analoger Größen unumgänglichen Vergrößerung absieht. Aber es ist die Frage, ob die digitale Lösung bei Problemen sinnvoll ist, für die es, wie etwa für die Verstärkung analoger Signale oder die Erzeugung elektrischer Schwingungen, gute analoge Schaltungen gibt. Die Antwort hängt wesentlich vom Stand der technischen Realisierung von Digital-Schaltungen in der Mikroelek-

tronik ab, d. h. davon, wieviele Bauelemente in einem Schaltkreis integriert werden können, wie schnell der Schaltkreis ist und wieviel er kostet. Es gibt Anzeichen dafür, daß mit fortschreitender Technik die analogen Schaltungen immer mehr durch digitale ersetzt werden. Die analoge Form der Informationsverarbeitung wird dadurch sicher nicht gänzlich verschwinden, zumal die Mikroelektronik auch bei den Analogschaltungen zu wesentlichen Verbesserungen führt. Aber sie könnte auf wenige spezielle Probleme verwiesen werden, für die sich die digitale Lösung nur schlecht eignet.

Wir schließen dieses Kapitel über elektronische Schaltungen für die Informationsverarbeitung mit einer kurzen Betrachtung ab, die für die technische Realisierung dieser Schaltungen von großer Bedeutung ist. Es geht um die Anzahl der Bauelemente. Um zu einer ungefähren Vorstellung zu kommen, wieviele Bauelemente in einer bestimmten Schaltung benötigt werden, wollen wir die für eine elementare logische Operation, d. h. die für ein Logik-Gatter erforderliche Zahl von Bauelementen, zu 10 annehmen. Für den Zählschaltkreis in Abb. 31 benötigt man sechs Logik-Gatter, also etwa 60 Bauelemente. In einer Quarzuhr werden etwa 30 von diesen Schaltkreisen benötigt, also etwa 2000 Bauelemente. Der Volladder in Abb. 36c mit 13 Logik-Gattern führt auf etwa 130 Bauelemente. Für die dynamische Speicherzelle in Abb. 28 werden zwei Bauelemente benötigt. Bei einem Speicher für 10^5 Bit ergeben sich daraus über 200 000 Bauelemente. Natürlich stellen diese Zahlen

nur sehr grobe Abschätzungen dar, die ohne weiteres um den Faktor 2 bis 3 von den wirklichen Werten abweichen können. Sie machen aber deutlich, um welche Größenordnung es bei der Anzahl von Bauelementen geht, die in Schaltungen für die Informationsverarbeitung benötigt werden.

Im nächsten Kapitel werden wir die Frage behandeln, wie sich Schaltungen mit einer so hohen Anzahl von Bauelementen industriell herstellen lassen.

5. Herstellungstechnologie der Mikroelektronik

Eine elektronische Schaltung, wie wir sie in Kapitel 4 besprochen haben, besteht aus Transistoren, Dioden, Kondensatoren und Widerständen. Wenn wir wissen wollen, wie eine solche Schaltung industriell hergestellt wird, werden wir uns also vor allem mit der Herstellung dieser Bauelemente zu beschäftigen haben. So jedenfalls müßte man an die Behandlung der Herstellungstechnologie für mikroelektronische Schaltungen herangehen, wenn diese in der Industrie tatsächlich aus einzelnen Bauelementen zusammengesetzt würden, ähnlich wie das ein Elektronik-Bastler tut, der sich einzelne Bauelemente kauft und diese dann durch Drähte zu einer Schaltung verbindet. Aber gerade so geht die Mikroelektronik-Industrie nicht vor.

Integrierte Schaltkreise

Die mikroelektronischen Schaltungen werden nicht durch nachträgliches Verbinden von vorher einzeln angefertigten Bauelementen hergestellt. Das Verbinden der Bauelemente erfolgt vielmehr gleichzeitig mit ihrer Herstellung. Es werden überhaupt nicht einzelne Bauelemente, sondern *Schaltungen im Ganzen* produziert. Man spricht deshalb von *integrierten Schaltungen* oder *Schaltkreisen*. Mikroelektronische Schaltungen sind integrierte, d. h. im Ganzen hergestellte Schaltungen. Die Herstellungstechnologie der

Mikroelektronik unterscheidet sich damit von vielen anderen industriellen Technologien, die nach dem Schema arbeiten «Herstellung von Teilen — Montage von Teilen». Gewisse Montage-Schritte sind zwar auch in der Mikroelektronik durchzuführen — das Halbleiterplättchen mit der integrierten Schaltung muß auf einen Träger mit den äußeren Anschlußpolen aufgebracht und mit diesen leitend verbunden werden. Schließlich muß das Ganze mit einem Gehäuse verkappt werden. Erst dann ist das Endprodukt, der mikroelektronische Schaltkreis, fertig. Aber das eigentliche Herzstück dieses Schaltkreises, die integrierte Schaltung auf dem Siliziumplättchen, der sogenannte *Chip* bzw. *Mikrochip*, wird im Ganzen, ohne die Montage von Teilen, hergestellt. Man spricht von einer integrierten Herstellungstechnik. Die Anwendung einer integrierten Technik bei der Herstellung von Autos würde etwa bedeuten, daß das Chassis als Ganzes aus einem Metallblock herausgearbeitet werden müßte und die Montage auf das Anbringen der Räder und das Verkleiden mit der Karosserie beschränkt bliebe. Abgesehen davon, daß ein solches Vorgehen eine technologische Unmöglichkeit wäre, besteht dafür in der Autoindustrie überhaupt keine Notwendigkeit. In der Mikroelektronik ist die integrierte Technik gleicher-

maßen vorteilhaft wie notwendig und möglich. Vorteilhaft für die Herstellung ist sie, weil die Komplettierung der Einzelbauelemente durch äußere Anschlußpole und ein schützendes Gehäuse ebenso entfällt wie das nachträgliche gegenseitige Verbinden der Bauelemente durch elektrische Leitungen. Die Bauelemente sind bereits verbunden. Anschlußpole und ein Gehäuse benötigt nur der gesamte Schaltkreis. Integrierte Schaltungen lassen sich deshalb billiger herstellen.

Die integrierte Herstellungstechnik bedeutet aber mehr als die Ausnutzung möglicher Kostenvorteile. Eine solche Technik ist notwendig, um Schaltungen mit 10000 und 100000 Bauelementen überhaupt industriell herstellen zu können. Dafür gibt es mehrere Gründe. Einer davon ist die Zuverlässigkeit von Schaltungen mit einer so hohen Anzahl von Bauelementen. Die Zuverlässigkeit ist viel zu gering bzw. die Ausfallrate viel zu hoch, wenn eine solche Schaltung durch nachträgliches Verbinden von Einzelbauelementen realisiert wird. Nehmen wir an, daß ein Einzelbauelement im Durchschnitt 10^6 Stunden, das sind immerhin über 100 Jahre, fehlerfrei funktioniert und dann ausfällt. Die Ausfallrate einer Schaltung aus 10000 solcher Bauelemente ist dann nach den Regeln der Wahrscheinlichkeitsrech-

nung ungefähr gleich $10^{-6} \cdot 10^4$ pro Stunde = 10^{-2} pro Stunde. Man muß also erwarten, daß die Schaltung bereits nach 100 Stunden nicht mehr funktioniert. Das ist unakzeptabel wenig. Da eine integrierte Schaltung viel Ähnlichkeit mit einem einzelnen, nur eben besonders kompliziert aufgebauten Bauelement hat, darf man hoffen, daß die mittels integrierter Technik hergestellten Schaltungen die hohe Zuverlässigkeit der Einzelbauelemente wenn schon nicht ganz erreichen, so doch auch nicht um das 10^4 -fache unterschreiten werden wie in unserem Beispiel, insgesamt also wesentlich zuverlässiger sein werden als die diskret aufgebauten Schaltungen. Diese Erwartung bestätigt sich voll auf. Integrierte Schaltungen *sind* zuverlässiger als Schaltungen aus Einzelbauelementen, und zwar nicht fünf- oder zehnfach, sondern 1000- und 10000mal zuverlässiger. Erst dadurch werden Schaltungen mit 10- und 100tausend Bauelementen, die eine ausreichende Zuverlässigkeit besitzen, überhaupt möglich. Andere Gründe, die den Übergang zu einer integrierten Herstellungstechnik notwendig machen, hängen mit der Größe der Bauelemente zusammen. Würde man für den Aufbau einer Schaltung Bauelemente von $0,1 \text{ cm}^3$ Größe und $0,1 \text{ g}$ Gewicht nehmen, so wäre das Volumen einer Schaltung mit 10000 Bauelementen größer als ein Litermaß und das Gewicht schwerer als die darin enthaltene Wassermenge. Das mag in manchen Fällen tolerierbar sein, in anderen sicher nicht. Man wird zu kleineren Bauelementen übergehen müssen. Der Zwang zur Verkleinerung der Bauelemente resultiert aber

nicht nur und nicht einmal in erster Linie aus der Forderung nach kleinen und leichten Schaltkreisen, also nach guten geometrischen und mechanischen Eigenschaften. Entscheidender ist, daß auch gute elektrische Eigenschaften, vor allem eine hohe Schaltgeschwindigkeit und ein geringer Leistungsverbrauch, nur mit kleinen Bauelementen möglich sind. Die Kleinheit der Bauelemente berührt also die Substanz der Schaltung, nämlich ihr *elektrisches* Verhalten. Es wird besser, wenn die Bauelemente kleiner werden. Um die für die Informationsverarbeitung wünschenswerten elektrischen Eigenschaften erzielen zu können, müssen Abmessungen im Bereich einiger hundertstel Millimeter erreicht und unterschritten werden. Bauelemente von einigen hundertstel Millimeter Länge, Breite und Höhe können einzeln, wenn überhaupt, aber nur mit sehr hohem technischen Aufwand und zu einem sehr hohen Preis hergestellt und mit Drähtchen untereinander verbunden werden. Praktisch kommt eine andere als die integrierte Herstellungstechnik für Schaltungen mit derart kleinen Bauelementen überhaupt nicht in Frage. Das Ergebnis unserer Überlegungen zur Herstellung von Schaltungen für die Informationsverarbeitung können wir also so zusammenfassen: Nur durch eine integrierte Technik lassen sich so zuverlässige, platzsparende, leichte, schnelle, energiesparende und billige Schaltungen herstellen, wie sie die Informationstechnik benötigt. Daß es aber auch möglich ist, eine integrierte Herstellungstechnik anzuwenden, hängt eng mit der Beschaffenheit der Halbleiterbauele-

mente und der aus ihnen bestehenden Schaltungen zusammen. Vergewärtigen wir uns noch einmal den Aufbau eines solchen Halbleiterbauelementes, etwa den des MOS-FET, so wie er in Abb. 16 dargestellt ist. Der MOS-FET stellt rein äußerlich ein Siliziumkristall-Plättchen dar. In diesem befindet sich ein p-dotierter Siliziumbereich, auf den eine Isolatorschicht aus Siliziumdioxid aufgebracht ist, die ihrerseits von einer Aluminium-Metallschicht bedeckt wird. Außerdem sind im Kristallplättchen zwei n-dotierte Siliziumbereiche untergebracht, die beide mit einer Aluminium-Metallschicht überzogen sind. Die räumliche Abmessung des MOS-FET in der horizontalen Richtung beträgt einige $10 \mu\text{m}$. Dazu liefern die beiden n-Gebiete rechts und links den Hauptbeitrag, das mittlere p-Gebiet, der Kanal, muß, wie wir aus Kapitel 3 wissen, möglichst schmal gemacht werden, damit der MOS-FET möglichst schnell wird. Typische Kanal-längen liegen heute bei etwa $3 \mu\text{m}$, aber auch bereits darunter. In der vertikalen Richtung ist die Ausdehnung des FET mit einigen Mikrometern relativ gering. Besonders die Oxidschicht muß sehr dünn gehalten werden (unter $0,1 \mu\text{m}$), um die Leitfähigkeit des Kanals über das Gate gut steuern zu können. Wenn wir bedenken, daß das Siliziumkristallplättchen für den MOS-FET wenigstens einige Zehntel Millimeter dick gemacht werden muß, damit es nicht bei der leisesten Berührung zerbricht, so wird deutlich, daß der FET nur eine relativ dünne Schicht von etwa 1 % Gesamtdicke unter der Oberfläche des Kristallplättchens beansprucht. Die 99 % des darunterlie-

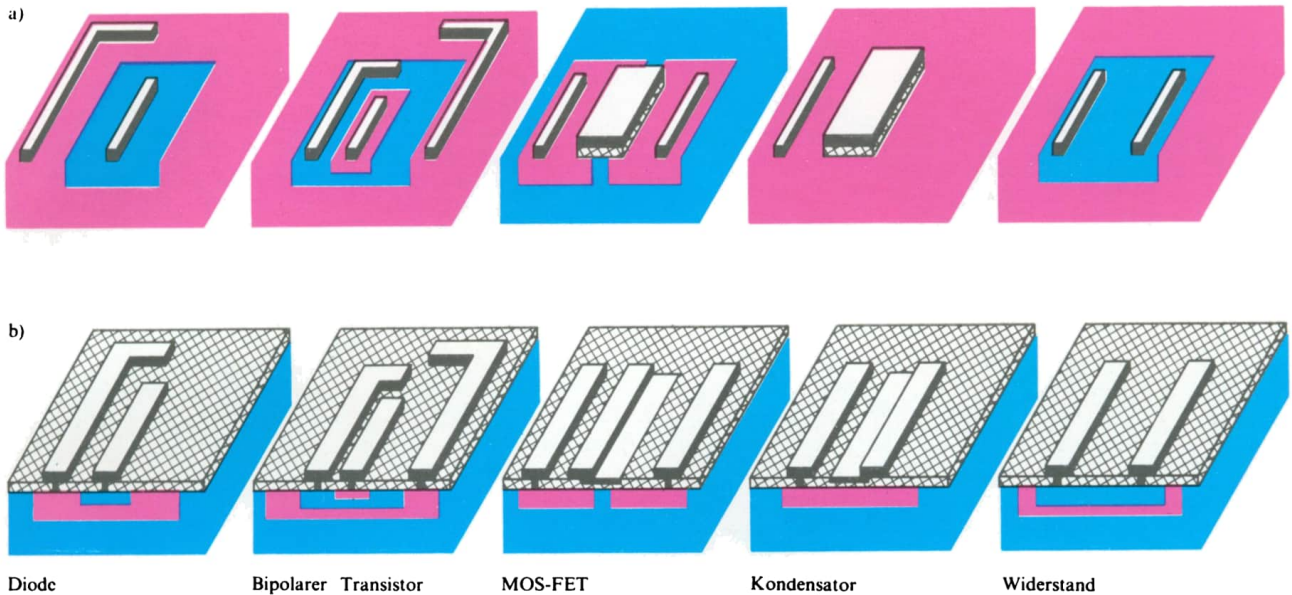


Abb. 38 Realisierungsmöglichkeiten für Halbleiterbauelemente an der Oberfläche eines Silizium-Kristallplättchens:

a) einzelne Bauelemente,

b) die gleichen Bauelemente wie darüber, jetzt aber mit einer Oxidschicht an der Oberfläche und einem zusätzlichen pn-Übergang zur gegenseitigen Isolation innerhalb eines Schaltkreises

genden Siliziummaterials haben nur die Funktion eines Trägers für den MOS-FET. Der Aufbau eines MOS-FET läßt sich allgemein also folgendermaßen charakterisieren: Er ist eine räumliche Neben- und Übereinanderanordnung von Bereichen aus unterschiedlichen Materialien dicht unterhalb der Oberfläche eines Silizium-Kristallplättchens. Die verschiedenen Materialien sind p-dotiertes Silizium, n-dotiertes Silizium, Siliziumdioxid und Aluminium. Ähnlich wie der MOS-FET lassen sich auch die anderen Halbleiterbauelemente beschreiben. Auch Bipolartransistoren, Dioden, Kondensatoren und Widerstände sind nichts weiter als über- und nebeneinander angeordnete Bereiche aus den genannten Materialien an der Oberfläche

eines Siliziumplättchens. An Hand von Abb. 38 kann man sich davon unmittelbar überzeugen. Wir wollen jetzt aus den Halbleiterbauelementen in Abb. 38 in Gedanken eine Schaltung aufbauen, indem wir die einzelnen Bauelemente-Plättchen auf einer Glasplatte nebeneinander anordnen und mit Aluminium-Drähten untereinander sowie mit den äußeren Anschlußpolen verbinden. Daß dies wegen der Kleinheit der Plättchen nur in Gedanken möglich ist, haben wir weiter oben bereits betont. In Abb. 39 ist der auf diese Weise erzeugte NAND-Schaltkreis aus Abb. 24b dargestellt. Sehen wir einmal von den Luftspalten zwischen den einzelnen Plättchen ab und denken uns die Aluminium-Drähtchen auf die

vorher noch mit einer isolierenden Oxidschicht überzogenen Oberflächen der Plättchen aufgeklebt, so ist auf der Glasplatte ein Gebilde entstanden, auf das voll die Beschreibung für ein einzelnes Bauelement zutrifft: Die Schaltung ist eine räumliche Neben- und Übereinanderanordnung von Materialbereichen aus p- und n-Silizium, Siliziumdioxid und Aluminium, und zwar an der Oberfläche eines Siliziumplättchens. Der einzige Unterschied ist, daß das Plättchen größer und die Anzahl der verschiedenen Materialbereiche höher ist als beim einzelnen Bauelement. Abgesehen von der Größe und Kompliziertheit ist aber der Aufbau der Schaltung der gleiche wie beim einzelnen Bauelement. Mit der gleichen Technologie, mit der sich Bau-

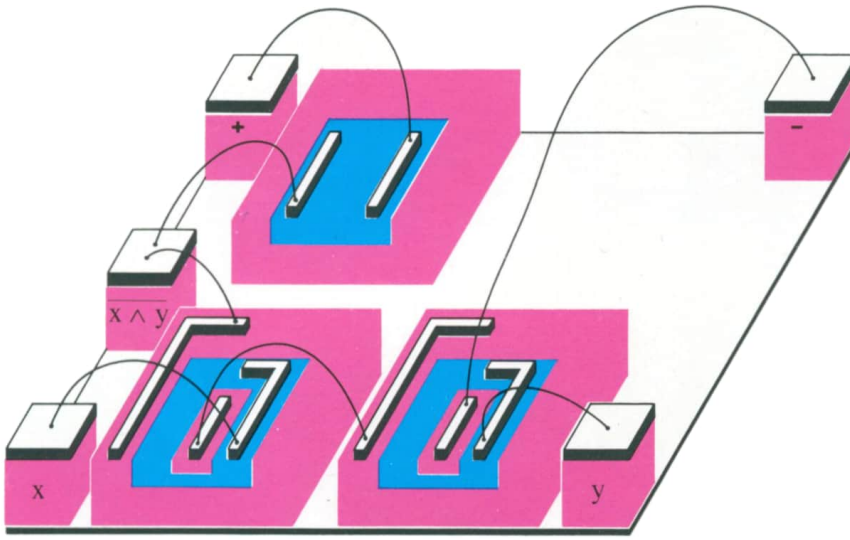


Abb. 39 Aufbau einer Schaltung aus Halbleiterbauelementen

elemente herstellen lassen, können daher auch ganze Schaltungen hergestellt werden. Das hatten wir konkret im Auge, als wir weiter oben davon sprachen, daß die besondere Beschaffenheit der Halbleiterbauelemente und der aus ihnen aufgebauten Schaltungen in der Mikroelektronik etwas möglich macht, was der Autoindustrie versagt ist — nämlich ihr Produkt im Ganzen, d. h. mit Hilfe einer integrierten Technik, herzustellen.

Auf eine Schwierigkeit bei der integrierten Herstellung von Schaltungen muß allerdings noch eingegangen werden. Sie hängt mit den Luftspalten zwischen den einzelnen Bauelemente-Plättchen in Abb. 39 zusammen, von denen wir zunächst abgesehen hatten. Diese Spalte haben aber eine wichtige Funktion. Sie isolieren die einzelnen Bauelemente voneinander. Würde man die Plättchen ohne weitere Vorsichtsmaß-

nahmen miteinander in Kontakt bringen, so wären die Bauelemente an Stellen miteinander leitend verbunden, wo es die Schaltung verbietet. Die Schaltung würde nicht funktionieren. Sollen die Luftspalte wegfallen, so muß ihre isolierende Funktion auf andere Weise realisiert werden. Zu diesem Zweck wird zwischen den einzelnen Bauelementen meist ein in Sperrichtung gepolter pn-Übergang angebracht, sofern nicht ohnehin einer vorhanden ist. Letzteres ist nur beim MOS-FET der Fall, bei den anderen Bauelementen in Abb. 38a jedoch nicht. Diese müssen jeweils noch in ein p-Gebiet eingebettet werden, so wie das in Abb. 38b dargestellt ist. Dadurch erhöht sich zwar der Platzbedarf für die Bauelemente — es deutet sich hier schon an, daß für die Integration *sehr vieler* Bauelemente in einem Schaltkreis MOS-FETs besonders gut geeignet sein werden —, aber der prinzipielle Aufbau der Bau-

elemente bleibt der gleiche, es muß nur noch ein weiterer Materialbereich aus p-Silizium hinzugefügt werden. Auch zwei Schichten aus Siliziumdioxid, die zur Isolation der Verbindungsleitungen und zu deren Schutz gegen äußere Einflüsse aufgebracht werden müssen, ändern am prinzipiellen Aufbau der Bauelemente nichts. In Abb. 38b ist die untere dieser Schichten mitgezeichnet.

Planartechnik

Die konkrete Ausgestaltung der integrierten Technik zur Herstellung mikroelektronischer Schaltkreise ist wesentlich dadurch geprägt, daß die Schaltkreise an der Oberfläche der Siliziumplättchen untergebracht sind und daß in vertikaler Richtung nur einige wenige verschiedene Materialschichten übereinanderliegen (im Unterschied zur horizontalen Richtung, wo sehr viel mehr und wesentlich ausgedehntere Materialbereiche aufeinanderfolgen). Dadurch ergibt sich die Möglichkeit, den Schaltkreis durch Bearbeitung der Oberfläche des Siliziumplättchens zu erzeugen. Bearbeitungsschritte, die in größere Tiefe gehen oder die von der Seite und von unten her auf das Werkstück einwirken, wie sie z. B. in der metallverarbeitenden Industrie ständig vorkommen, sind in der Mikroelektronik kaum erforderlich. Von der Bearbeitung ist im wesentlichen nur eine einzige Ebene des Werkstückes «Siliziumplättchen» betroffen, nämlich die obenaufliegende Oberfläche. Man bezeichnet ein solches Bearbeitungsverfahren als *Planartechnik*. Das Werkstück, das, wie wir später noch sehen werden, gar kein einzelnes Plättchen, sondern eine größere Siliziumscheibe ist, die man in sehr viele Plättchen zerteilen kann, nennt



Abb. 42 Arbeit im Reinraum: Beschickung eines Diffusionsofens mit Siliziumscheiben

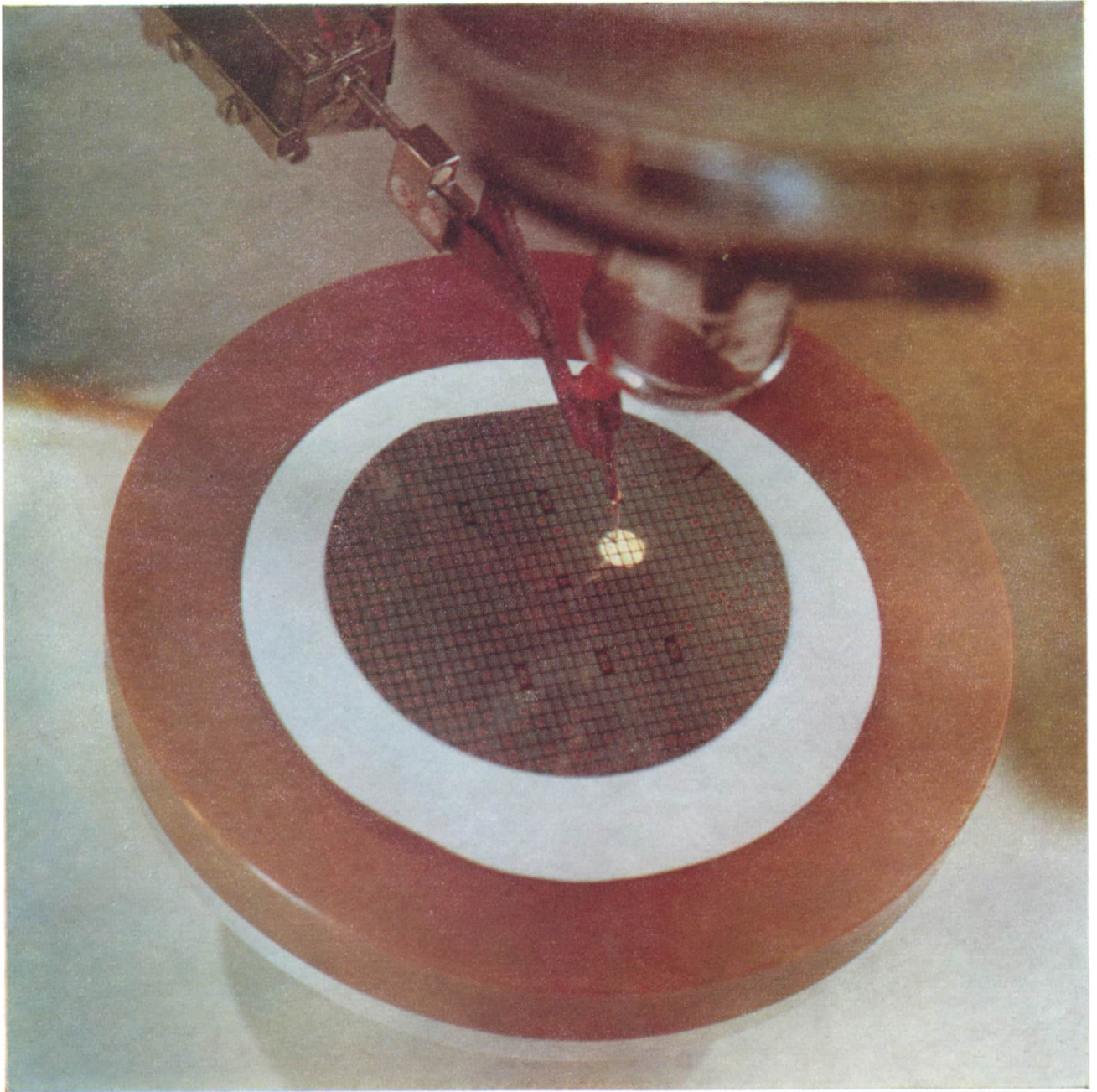


Abb. 43 Siliziumsubstratscheibe mit angeritzten Plättchen

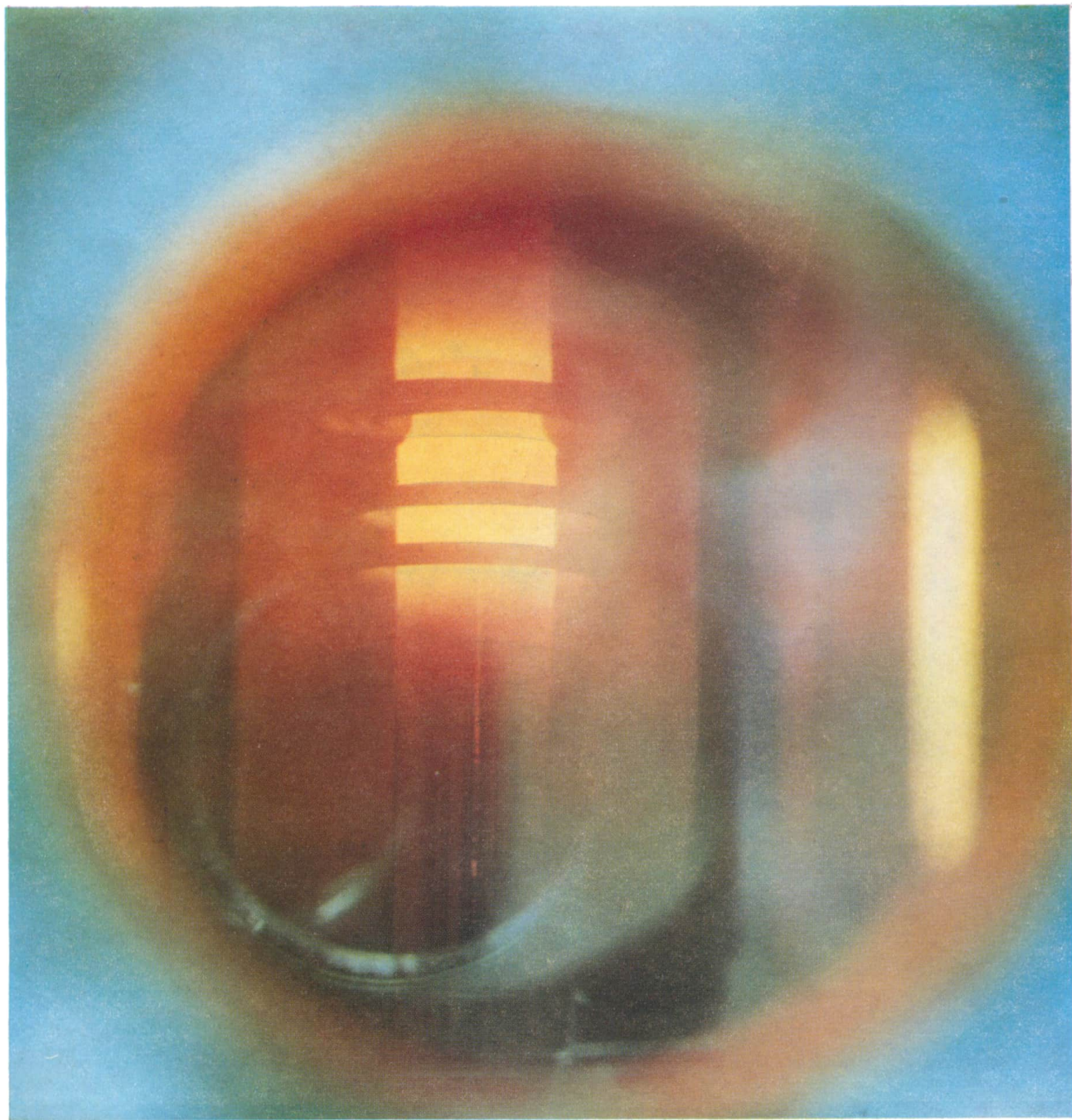


Abb. 50 Züchtung von Silizium-Einkristallstäben nach dem Zonenschmelzverfahren. Die dunklen Ringe in der glühenden Schmelzzone des Siliziumstabes stammen von Heizspulen.

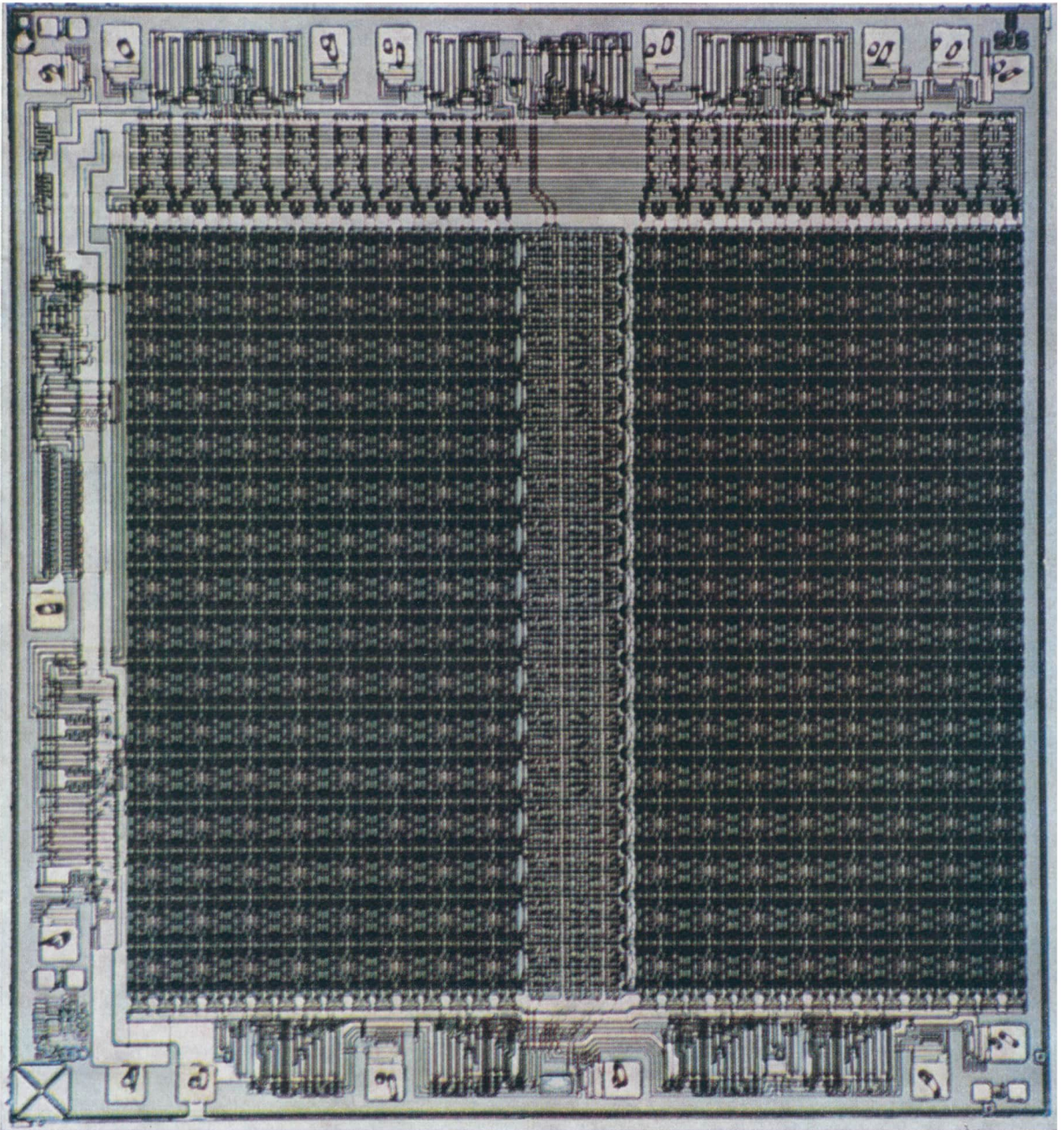


Abb. 45 Chipaufnahme des Speicherschaltkreises U202 (202)

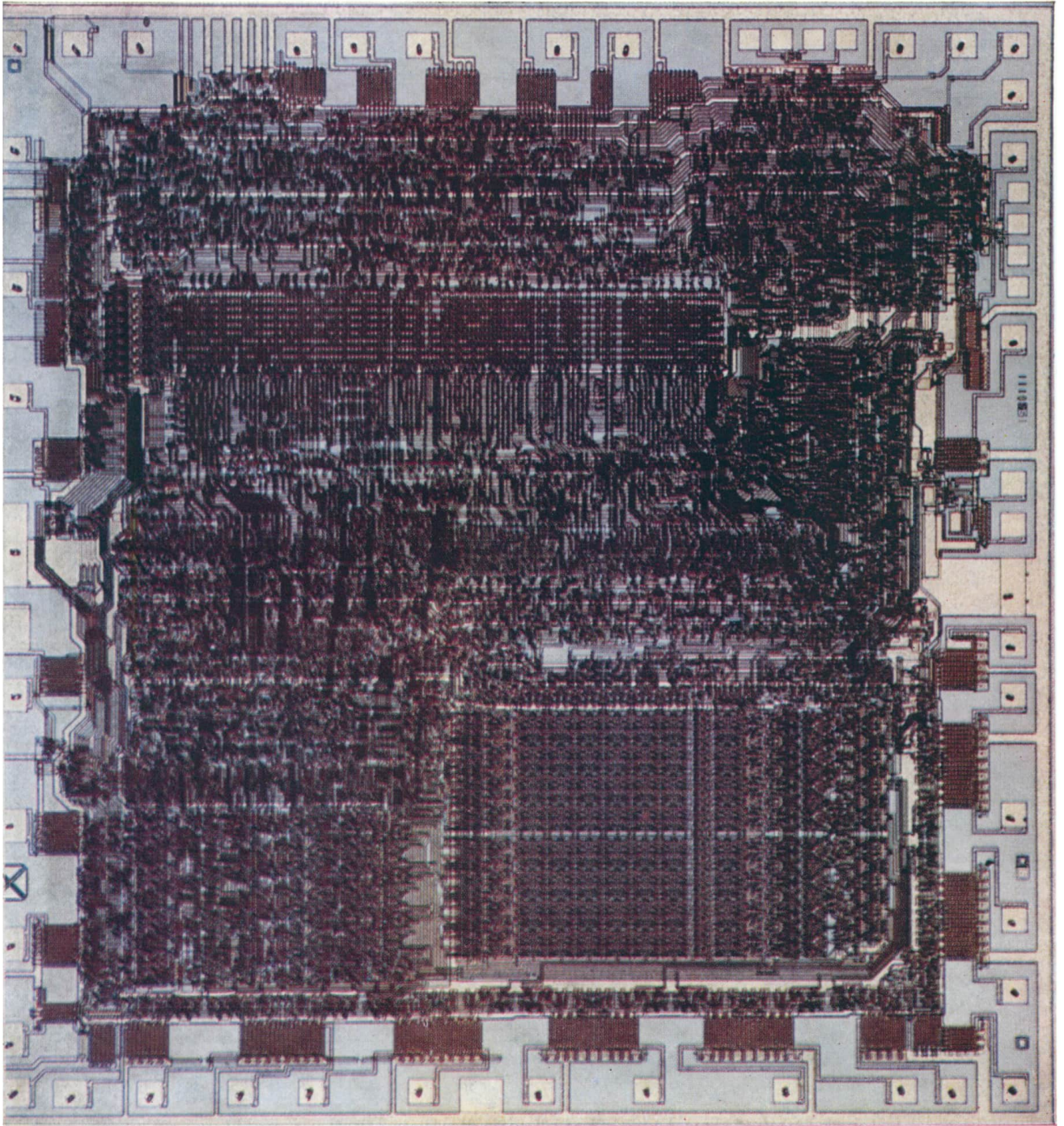


Abb. 46 Chipaufnahme des Mikroprozessorschaltkreises U880 (Z80)

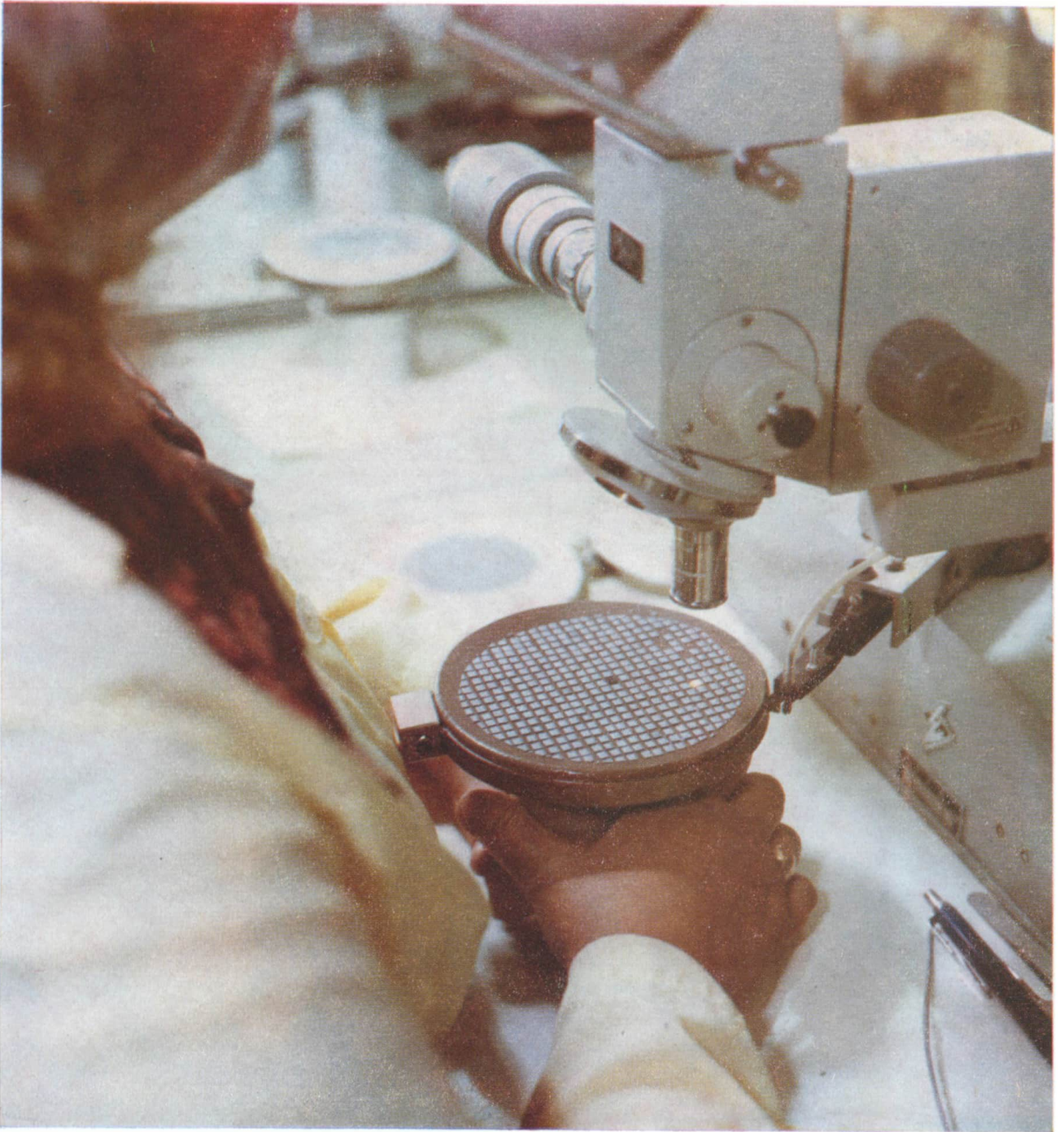


Abb. 51 Zerteilen einer Siliziumscheibe in einzelne Chips

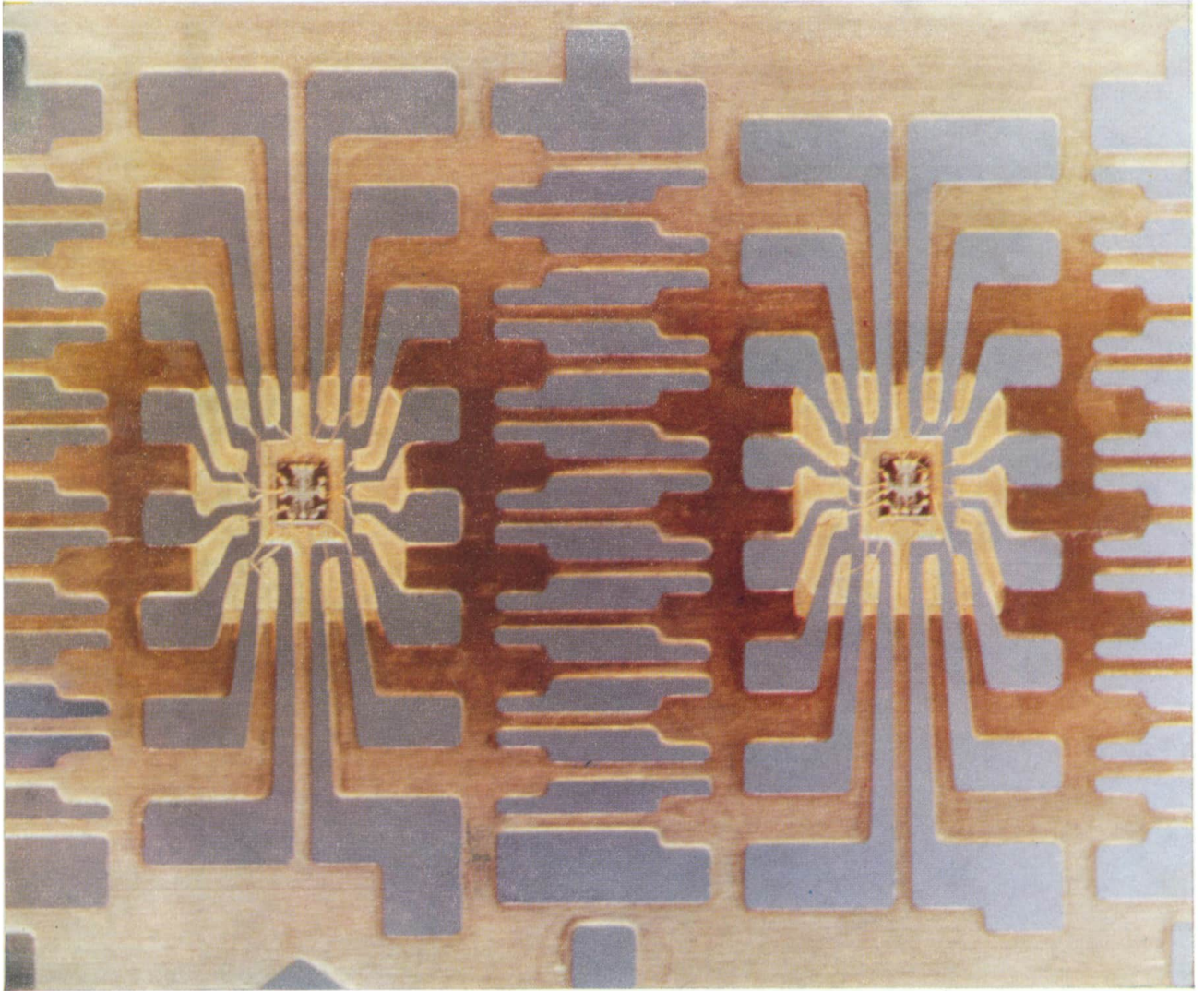


Abb. 52 Chips auf einem metallischen Trägerstreifen. Das Bonden ist bereits ausgeführt. Die Bondinseln der Chips sind durch feine Golddrähtchen mit den Fingern für die äußeren Anschlußkontakte verbunden.

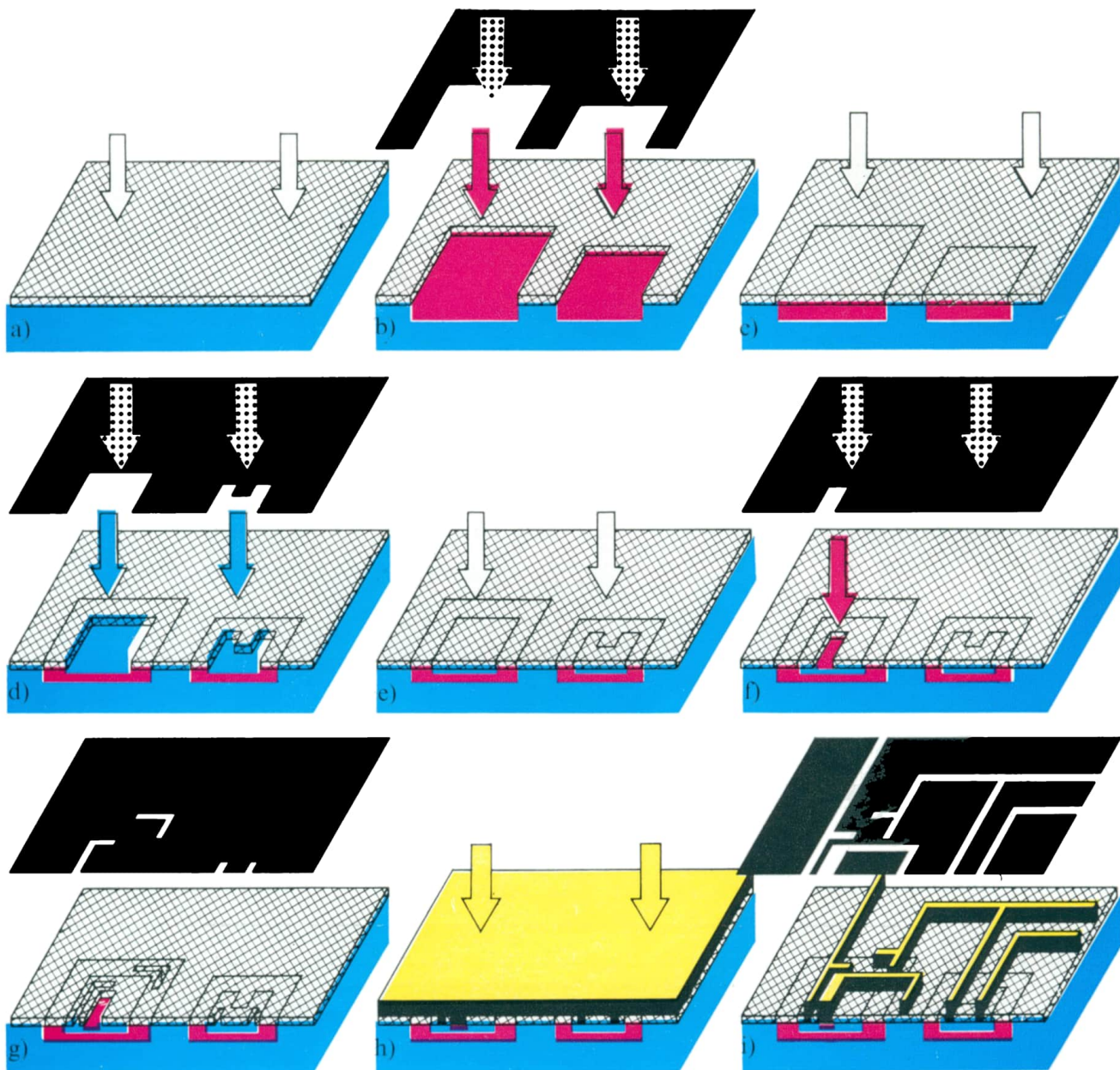


Abb. 41 Prinzip der Herstellung eines integrierten Schaltkreises nach der Silizium-Planartechnik

man *Substrat*. Die Herstellung integrierter Schaltkreise erfolgt also mit Hilfe einer Planartechnik, die auf das Siliziumsubstrat angewandt wird. Im folgenden soll auf diese Technik näher eingegangen werden. Wir beginnen mit dem zentralen Teil dieser Technik, dem sogenannten Zyklus 1, und behandeln die vor- und nachgelagerten Schritte, die man mitunter als Zyklus 0 und Zyklus 2 bezeichnet, im Anschluß.

Zyklus 1: Oxydation

Ausgangspunkt des Zyklus 1 ist das noch unberührte Siliziumsubstrat. Zunächst wird an seiner Oberfläche durch Einwirkung von Sauerstoff bei einer Temperatur von etwa 1000 °C eine Siliziumdioxidschicht erzeugt. Der Sauerstoff wird dabei meist mit Wasserdampf gemischt, um die Oxydation zu beschleunigen. Durch Steuerung der Einwirkungszeit des Sauerstoffs und der Temperatur des Substrats kann die Dicke der Oxidschicht leicht eingestellt werden. In trockenem Sauerstoff wächst z. B. in einer Stunde bei 1050 °C eine Oxidschicht von 0,1 µm Dicke. Bei Zugabe von Wasserdampf wird die Schicht unter sonst gleichen Bedingungen 0,5 µm dick.

Das Siliziumdioxid ist ein chemisch sehr beständiges, mechanisch festes und elektrisch gut isolierendes Material, das auf dem darunterliegenden Silizium sehr gut haftet. Bei der Herstellung integrierter Schaltkreise dient die Siliziumdioxidschicht außer zu *Isolation*szwecken vor allem zur *Passivierung*, d. h. zur elektrischen und chemischen Stabilisierung der Siliziumoberfläche, sowie zur *Maskierung* des Substrats gegen unerwünschte Einflüsse aus den nachfolgenden Bearbeitungsschritten. Die sehr gute

Eignung des Siliziumdioxids für Maskierungs-, Passivierungs- und elektrische Isolationsschichten ist, neben den in Kapitel 3 genannten physikalischen Gesichtspunkten, der ausschlaggebende Grund dafür, daß es unter allen Halbleitermaterialien gerade das Silizium ist, das in der Mikroelektronik angewandt wird. Sollen bestimmte Bereiche des unter der Oxidschicht liegenden Siliziums einer Bearbeitung, etwa einer Dotierung mit Phosphoratomten, unterzogen werden, so muß an diesen Stellen vorher die Oxidschicht entfernt werden. Die Markierung dieser

Strukturierung

durch Fotolithographie

Stellen auf dem Siliziumsubstrat, die sogenannte Strukturierung, erfolgt mit Hilfe eines Verfahrens, das man als *Fotolithographie* bezeichnet. «Fotolithographie» hat etwas mit Licht (von *Foto*), Stein (von *Lithos*) und Schreiben (von *Graphie*) zu tun: Mit Hilfe von Licht werden in die Oxidschicht die gewünschten flächenhaften Strukturen eingeschrieben. Dazu wird die gesamte Oxidschicht zunächst mit einem lichtempfindlichen Fotolack überzogen. Die Fotolackschicht wirkt wie ein fotografischer Film. Man kann sie belichten und entwickeln. Von den belichteten Stellen kann der Fotolack mit Hilfe eines organischen Lösungsmittels entfernt werden. An den freigelegten Stellen, aber auch nur an diesen, wird bei der anschließenden Behandlung der Oberfläche mit einer Säure das Siliziumdioxid aufgelöst. Läßt man die Säure lange genug einwirken, tritt das darunterliegende Silizium zutage und ist wie durch ein Fenster für den folgenden Bearbeitungsschritt zugänglich. Man spricht

Ätzen

vom *Ätzen des Siliziumdioxids*.

Neben dem Ätzen mittels flüssiger Chemikalien (sogenanntes naßchemisches Ätzen) wird heute in zunehmendem Maße das sogenannte Plasmaätzen eingesetzt. Dazu werden halogenhaltige Gase verwendet. Das Gas befindet sich zusammen mit dem zu ätzenden Substrat in einer Reaktionskammer. Dort wird eine elektrische Gas- (oder Plasma-) Entladung gezündet, bei der die Halogenverbindung in ihre Bestandteile zerfällt. Darunter sind auch stark reaktive Halogenionen, die das Siliziumdioxid zersetzen. Wird die Reaktionskammer mit anderen Gasen gefüllt, so kann statt der Siliziumdioxidschicht auch eine Silizium- oder Metallschicht weggeätzt werden. Auch die Beseitigung des Fotolacks mittels Plasmaätzen ist möglich. Da bei der Herstellung eines Schaltkreises mehrere Bearbeitungsschritte, z. B. Dotierung mit Phosphor, anschließend mit Bor, hintereinander auszuführen sind, muß die Fotolithographie mehrfach angewendet werden. Vorher muß die Oberfläche jedesmal wieder durch Oxidation mit einer durchgehenden Siliziumdioxidschicht überzogen werden. Danach müssen die für den nächsten Bearbeitungsschritt freizulegenden Bereiche der Oberfläche belichtet werden. Dazu kann man das Prinzip der Diaprojektion benutzen. Das Dia oder die Fotoschablone, wie man gewöhnlich sagt, enthält die zu belichtenden Bereiche als helle, die übrigen als dunkle Stellen. Die Belichtung erfolgt wegen der größeren Schärfe der Konturen mit möglichst kurzwelligem Licht. Neben dem hier beschriebenen positi-

ven Fotolack, bei dem die belichteten Stellen aufgelöst werden und die dunklen stehen bleiben, gibt es auch einen negativen Fotolack, bei dem genau das Umgekehrte der Fall ist.

Dotierung

An die Fotolithografie schließt sich die Dotierung mit Phosphor oder Arsen zur Erzeugung von n-leitendem Silizium einer gewünschten Ladungsträgerkonzentration oder mit Bor zur Erzeugung von p-Silizium an. Dazu wird entweder ein Diffusionsverfahren angewandt, d. h. bei 900 bis 1200 °C wird die Oberfläche des Substrats den Dotierungsatomen ausgesetzt, die dann in das Substrat hineindiffundieren. Oder man benutzt die Ionenimplantation. Beim zuletzt genannten Verfahren schießt man elektrisch geladene Phosphor-, Arsen- oder Boratome, also Ionen dieser Elemente, die in einem elektrischen Feld beschleunigt wurden, auf die Substratoberfläche auf.

Wie Geschosse in eine Wand dringen die Ionen vermöge ihrer Bewegungsenergie ein Stück in das Siliziumsubstrat ein und kommen dort zum Stillstand. Sie werden gewissermaßen in das Silizium eingepflanzt oder implantiert. Auf ihrem Weg haben die Ionen den regelmäßigen Kristallaufbau sehr stark gestört. Die Schäden müssen anschließend durch eine Wärmebehandlung bei 500–900 °C oder durch kurzzeitiges starkes Aufheizen mittels Blitzlicht oder Laserstrahl wieder ausgeheilt werden. Sowohl die Diffusion als auch die Ionenimplantation werden nur an den vom Siliziumdioxid befreiten Stellen des Substrats wirksam. In der Oxidschicht bleiben die Dotierungsatome stecken.

Epitaxie

Kristalline Siliziumschichten, die in bestimmter Weise dotiert sind, können auf dem Siliziumsubstrat auch durch Abscheiden von außen erzeugt werden. Man spricht dann von *Epitaxie*. Bei einer gewöhnlichen Abscheidung, etwa beim Aufdampfen von Silizium auf eine Glasplatte, entsteht eine strukturell ungeordnete Schicht. Die Besonderheit der epitaktischen Abscheidung besteht darin, daß die Schicht einkristallin wird. Dafür sorgt die einkristalline Unterlage. Sie prägt der abgeschiedenen Schicht ihre eigene kristalline Struktur auf. In der Praxis verwendet man meist die Abscheidung aus der Gasphase (ein Gemisch von gasförmigem SiCl_4 , Wasserstoff und einem arsen- bzw. phosphorhaltigem Gas zur Dotierung werden über die erhitzte Siliziumunterlage geleitet), seltener die Abscheidung aus der flüssigen Phase (ein Schiffchen mit der Siliziumschmelze und eventuellen Dotierungsbestandteilen wird langsam über die Siliziumunterlage hinweggezogen). Es entsteht eine kristalline Siliziumschicht mit der gewünschten Dotierung. Durch die Epitaxie werden die Möglichkeiten für die vertikale Strukturierung der Bauelemente und Schaltkreise vielfältiger. Vor allem deshalb wird die Epitaxie neben den bereits genannten Verfahren zur Realisierung von dotierten Schichten zusätzlich eingesetzt.

Metallisierung

Zur Erzeugung der leitenden Verbindungen, der sogenannten Leitbahnen, zwischen den Bauelementen des Schaltkreises und zwischen diesen und den Kontaktstellen für die äußeren Anschlüsse, den sogenannten

Bondinseln, wird das Substrat vollständig mit einer Aluminiumschicht überzogen. Dazu wird Aluminium durch Aufheizen verdampft bzw. mittels eines Elektronen- oder Ionenstrahls zerstäubt. Auf dem kälteren Siliziumsubstrat scheidet sich das Aluminium als metallische Schicht ab. Die Metallschicht wird mit Hilfe der Fotolithografie weiter bearbeitet. Ähnlich wie früher an ausgewählten Stellen das Siliziumdioxid stehenblieb und an den übrigen Stellen durch Ätzen entfernt wurde, wird jetzt das Metall bis auf die Leitbahnen und die Bondinseln wieder aufgelöst (siehe Abb. 40, S. IX). Anschließend wird die gesamte Oberfläche mit einer Schutzschicht — meist aus Silikatglas — überzogen, die nur an den Bondinseln wieder entfernt wird.

Beispiel

An einem etwas hypothetischen Beispiel, der Herstellung eines Schaltkreises für die Negation, bestehend aus einem npn-Transistor und einem Widerstand, wollen wir uns den oben beschriebenen Herstellungsprozeß noch einmal vergegenwärtigen. So wie in diesem Beispiel wird heute kein Schaltkreis wirklich hergestellt, aber das Wesentliche wird bereits an diesem einfachen Beispiel deutlich. In Abb. 41 (siehe S. VIII) sind die einzelnen Herstellungsschritte dargestellt. Das Substrat, auf dem der Schaltkreis erzeugt werden soll, besteht aus p-Silizium. Durch Oxidation wird das p-Silizium mit einer Oxidschicht überzogen (Bild a). Nach Durchführung der Fotolithographie unter Verwendung der ersten Fotoschablone und Dotierung mit Phosphor entsteht Bild b, in dem bereits die Konturen des Transistors (links) und des Widerstandes (rechts)

sichtbar sind. Die Eindiffusion des Phosphors bei etwa 1200 °C führt automatisch zur Oxidation der gesamten Oberfläche und dient damit gleichzeitig der Vorbereitung des zweiten Fotolithographieschrittes (Bild c). Mit Hilfe einer zweiten Schablone werden zwei Siliziumbereiche freigelegt, die anschließend mit Bor p-dotiert werden (Bild d). Die Schrittfolge Oxidation (Bild e) — Fotolithographie — Dotierung (Bild f) wird noch einmal wiederholt, jetzt aber mit der dritten Schablone und mit Phosphor als Dotierungsmaterial. Es entsteht das zweite n-Gebiet des Transistors (Bild f). Nun werden unter Verwendung einer weiteren Schablone die Anschlußstellen für die Kontakte freigelegt (Bild g). Die Oberfläche wird ganz mit einer Aluminiumschicht überzogen (Bild h), aus der durch Fotolithographie die Anschlußpole für die beiden Bauelemente und die Leitbahnen für ihre gegenseitige Verbindung herausgearbeitet werden (Bild i). Dabei wurde im Unterschied zu den vorhergehenden Lithographieschritten, wo mit einem positiven Fotolack gearbeitet wurde, ein negativer Lack verwendet. Im Ergebnis der gesamten oben beschriebenen Schrittfolge entstehen ein npn-Transistor und ein Widerstand (Bild i), die jeweils durch einen pn-Übergang vom p-Substrat getrennt sind, falls letztere in Sperrrichtung gepolt werden. Dazu muß an das p-leitende Substrat eine negative Spannung angelegt werden und an die beiden angrenzenden n-Gebiete eine positive. Das äußere n-Gebiet des Transistors wird damit zum Kollektor. Das an den Kollektor angrenzende p-Gebiet ist die Basis und das innere n-

Gebiet der Emitter. Der Kollektor des Transistors ist mit dem Ausgang des Widerstandes verbunden. Der Eingang und der Ausgang des Widerstandes bilden zusammen mit dem Basis- und Emitterkontakt die vier äußeren Anschlüsse des Negationsschaltkreises in Abb. 41. Zur Herstellung dieses Schaltkreises mußten wir die Fotolithographie fünfmal, die Oxidation dreimal, die n-Dotierung zweimal, die p-Dotierung und Metallisierung jeweils einmal anwenden. Immerhin kein geringer Aufwand für einen so einfachen Schaltkreis. Aber etwas ganz Merkwürdiges ist der Fall: Wären auf den Fotoschablonen nicht nur Strukturen für einen Transistor und einen Widerstand, sondern für Zehntausende solcher Bauelemente einschließlich der notwendigen Leitbahnen und Bondinseln abgebildet gewesen, so wäre im Ergebnis derselben Prozeßschritte statt eines Schaltkreises mit zwei, ein Schaltkreis mit Zehntausenden von Bauelementen entstanden. Nicht ein einziger Prozeßschritt mehr hätte dafür durchgeführt werden müssen. Die Planartechnologie macht keinen Unterschied zwischen der Herstellung eines einzigen Bauelementes, eines Schaltkreises mit zwei oder eines Schaltkreises mit einer riesigen Anzahl von Bauelementen. Die Anzahl der Bauelemente in einem **Integrationsgrade** Schaltkreis bezeichnet man auch als dessen *Integrationsgrad*. Schaltkreise mit 100 Bauelementen oder darunter haben einen geringen Integrationsgrad — man spricht von SSI-Schaltkreisen (Small Scale Integration). Bei 100 bis 1000 Bauelementen spricht man von MSI-Schaltkreisen

(Medium Scale Integration), zwischen 1000 und einigen 10000 Bauelementen liegen die LSI-Schaltkreise (Large Scale Integration) und darüber die VLSI-Schaltkreise (Very Large Scale Integration). Gegenwärtig werden bereits VLSI-Schaltkreise mit mehreren 100000 Bauelementen hergestellt. Im Laboratorium sind Schaltkreise mit mehr als einer Million Bauelemente realisiert worden. Die Abmessungen eines Siliziumplättchens, auf dem ein Schaltkreis untergebracht ist, sind relativ klein. Nehmen wir für eine grobe Abschätzung an, daß ein Bauelement 20 Mikrometer lang und 20 Mikrometer breit ist, d. h. eine Fläche von $20\ \mu\text{m} \cdot 20\ \mu\text{m}$ beansprucht. Die Fläche zur Unterbringung von 40000 Bauelementen ist dann $4 \cdot 10^4 \cdot (2 \cdot 10^{-2}) \cdot (2 \cdot 10^{-2})\ \text{mm}^2 = 16\ \text{mm}^2$. Dazu kommen noch die Bondinseln. Ein quadratisches Plättchen muß also eine Seitenlänge von gut 4 mm haben, um einen Schaltkreis mit 40000 Bauelementen aufnehmen zu können. Das sind immer noch relativ kleine Abmessungen. Der Ausgangspunkt für die Herstellung der Schaltkreise sind Siliziumstäbe mit einem Durchmesser von ca. 70 mm, 120 mm oder 150 mm. Aus den Stäben erzeugt man zunächst dünne Scheiben, deren Dicke mit etwa 0,5 mm gerade noch so groß ist, daß sie nicht zerbrechen. Diese Substratscheiben werden aber nicht in Substratplättchen zerlegt, die dann einzeln weiter verarbeitet werden, wie man das vielleicht nach den bisherigen Ausführungen vermuten könnte. Eine Substratscheibe wird vielmehr als Ganzes den Prozeßschritten unterworfen, die wir am Substratplättchen erläutert haben. In Abb. 42 (siehe S. I) werden die

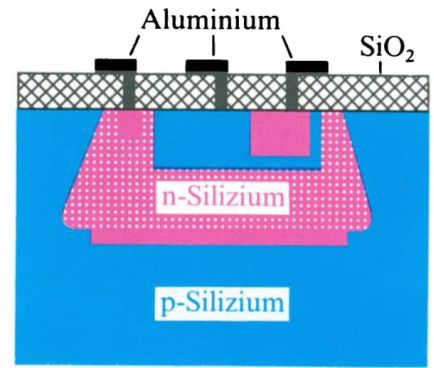
Scheiben gerade in den Diffusions-ofen eingeführt. Der Schaltkreis wird auf der Scheibe so oft erzeugt, wie es ihre Fläche hergibt. Die Integra ion auf der Ebene der Bauelemente wird gewissermaßen auf der Ebene der Schaltkreise fortgesetzt, allerdings nur bis zu einem bestimmten Punkt im Herstellungsprozeß, denn das Ziel sind ja *einzelne* Schaltkreise. Aus den eingeritzten Rechtecken auf der Substratscheibe in Abb. 43 (siehe S. II) werden später die Schaltkreislättchen. Man bezeichnet diese Plättchen auch als *Chips*. Auf einer Scheibe von 100 mm Durchmesser haben $\pi \cdot 50 \cdot 50 \text{ mm}^2$ dividiert durch $16 \text{ mm}^2 = 490$ Chips mit dem oben diskutierten Schaltkreis Platz. Man vergegenwärtige sich diese enorme Leistung: Im Ergebnis weniger Prozeßschritte werden auf einer Scheibe knapp 500 Schaltkreise mit je 40000 Bauelementen hergestellt. Die Siliziumplanartechnologie ist also ein außerordentlich produktives Verfahren zur Herstellung integrierter Schaltkreise. Das bleibt auch dann richtig, wenn man nur eine relativ geringe Ausbeute an einwandfrei funktionierenden Schaltkreisen in Rechnung stellt. Wir werden auf die Ausbeute noch genauer eingehen. Um die unbrauchbaren Schaltkreise zum frühest möglichen Zeitpunkt herauszufinden und zwecklose weitere Arbeiten mit ihnen zu vermeiden, prüft man alle Schaltkreise noch auf der Scheibe auf Herz und Nieren. Dazu werden auf die Bondinseln jedes Schaltkreises Kontaktspitzen aufgesetzt und nach einem bestimmten Testprogramm mit Spannungsimpulsen belegt. Daß es sich dabei um eine schwierige, zeitaufwendige Aufgabe handelt, wird

man ermessen, wenn man bedenkt, daß der bestandene Test das fehlerfreie Funktionieren jedes einzelnen Bauelementes und jeder Verbindung zwischen ihnen und den Bondinseln garantieren muß. Fehlerhafte Schaltkreise werden durch einen roten Tintenpunkt markiert (siehe Abb. 43), damit sie nach dem Zerteilen der Scheibe in die einzelnen Chips aussortiert werden können. In den Abb. 44—46 (siehe S. IV, V, X) zeigen wir mikroskopische Aufnahmen der Oberfläche einiger Chips aus der Produktion des Kombiniertes Mikroelektronik der DDR. Wie es unter der Oberfläche eines solchen Chips aussieht, ist in Abb. 47 (siehe S. IX) dargestellt.

Realisierungstechniken

Mit der Passivierung, der Fotolithographie, dem Ätzen, der Epitaxie, der Dotierung und der Metallisierung der Siliziumscheiben haben wir die wichtigsten Prozeßschritte zur Herstellung integrierter Schaltkreise kennengelernt. Damit ist geklärt, wie integrierte Schaltkreise aus Halbleiterbauelementen im Prinzip hergestellt werden können. Für die konkrete Realisierung des Herstellungsverfahrens sind aber noch viele Varianten möglich.

In Abb. 38 haben wir nur die einfachste Variante für den Aufbau von Halbleiterbauelementen aus verschiedenen Materialbereichen dargestellt. Die einfachste Variante ist aber in kaum einem Fall die beste. Bessere Bauelemente können durch andere Formgebung und Anordnung der einzelnen Materialbereiche und durch die Einfügung zusätzlicher Materialbereiche erzielt werden. Neben normal dotiertem n- und p-Silizium wird auch besonders stark



n⁺-Silizium

Abb. 48 Struktur eines Bipolar-Transistors, wie er heute in der Mikroelektronik wirklich verwendet wird.

dotiertes n⁺- und p⁺-Material eingesetzt. In Abb. 48 ist ein Beispiel für einen Bipolar-Transistor dargestellt, wie er heute wirklich in integrierten Schaltkreisen verwendet wird. Charakteristisch ist die zusätzliche vergrabene n⁺-Schicht. Bei unipolaren Transistoren in einem integrierten Schaltkreis können die Kanäle alle aus n- oder alle aus p-Material hergestellt werden, oder es können beide Kanalarten nebeneinander verwendet werden. Anstelle von Aluminium kann für die Realisierung der Gate-Elektrode eines MOS-FET hochleitendes polykristallines Silizium eingesetzt werden. Die gegenseitige Isolation der Bauelemente in einem Schaltkreis läßt sich nicht nur durch gesperrte pn-Übergänge, wie wir das beschrieben haben, sondern auch auf andere Weise realisieren. Diese und andere hier nicht genannte Möglichkeiten entsprechen den zahlreichen verschiedenen *Realisierungstechniken* der Mikroelektronik. In Tabelle 7 sind einige dieser Techniken aufge-

Tab. 7 Realisierungstechniken der Mikroelektronik

Bipolar-Techniken	MOS-Techniken
SBC (Standard Burried Collector)	Metall-Gate p- oder n-Kanal-MOS
ASBC (Advanced SBC)	Silizium-Gate p- oder n-Kanal-MOS
CDI (Collector Diffusion Isolation)	DMOS (Double Diffusion MOS)
BDI (Basis Diffusion Isolation)	VMOS (Vertical MOS)
OXIS (Oxid Isolation)	CMOS (Complementary MOS)
	SOS (Silicon On Saphir)

führt. Wir beschränken uns auf die Angabe der Bezeichnungen. Eine detailliertere Behandlung würde hier zu weit führen und außerdem nichts grundsätzlich Neues bringen.

Reinheit und Präzision

Auf einen Punkt, der alle Realisierungstechniken betrifft, müssen wir aber noch eingehen. Es handelt sich um die extrem hohen Anforderungen, die an die Reinheit aller verwendeten Materialien und Geräte und im Falle der Substratscheiben auch an die kristallographische Perfektion zu stellen sind. Diese hohen Anforderungen resultieren einerseits aus der extremen Kleinheit und Komplexität der verschiedenen Materialstrukturen in den Schaltkreisen, andererseits aber auch aus jener Eigentümlichkeit der Halbleiterkristalle, die wir in Kapitel 3 besprochen haben. Ihre elektrischen Eigenschaften können bereits durch geringste Mengen von Verunreinigungen und durch kleinste Strukturdefekte drastisch verändert werden. Der gesamte beschriebene Herstellungsprozeß muß deshalb mit Chemikalien von ungewöhnlich hoher Reinheit (99,999 %) und in speziell dafür hergerichteten, sogenannten Reinräumen, durchgeführt werden. Normales destilliertes Wasser

für die vielen Spülprozesse oder die gewöhnliche Zimmerluft sind viel zu unsauber und müssen extra gereinigt werden. Im Reinraum dürfen in einem Kubikmeter Luft nicht mehr als 3500 Staubteilchen mit einer Größe über 0,5 µm vorhanden sein. In der gewiß auch hochreinen Luft eines Operationssaales dürfen es immerhin 350000 sein. Das Bedienungspersonal für die Anlagen in den Reinräumen muß eine spezielle staubfreie Kleidung tragen. Die Kopfbedeckung muß die Haare fest umschließen. Abbildung 42 vermittelt einen Eindruck von der Arbeit in einem solchen Reinraum. Neben der Reinheit bei allen Verarbeitungsschritten müssen hohe Anforderungen an die geometrische Präzision der Fotolithographie gestellt werden. Man bedenke, daß Strukturen erzeugt werden müssen, die in ihren Abmessungen bis zu einem Mikrometer betragen können, wie etwa beim Kanal eines MOS-FET. Dazu muß die Genauigkeit der Projektion wenigstens eine Zehnerpotenz höher sein, die Ungenauigkeiten dürfen also einige Zehntel Mikrometer, das sind einige Zehntausendstel Millimeter, nicht überschreiten. Dazu müssen die Fotoschablonen genügend genau und die Schärfe der Abbil-

dung muß genügend hoch sein, vor allem müssen aber die Projektionen bei den nachfolgenden fotolithographischen Schritten exakt zu den Projektionen bei den vorhergehenden positioniert werden. Vergegenwärtigt man sich außerdem noch, daß die Substratscheiben mehrfach Temperaturen über 1000 °C ausgesetzt und wieder auf Zimmertemperatur abgekühlt werden müssen oder daß sie mit ätzenden Chemikalien in Berührung kommen, so versteht man, daß Material- und Strukturfehler in den Schaltkreisen nicht zu vermeiden sein werden. In vielen Fällen werden diese Fehler zur Funktionsuntüchtigkeit des Schaltkreises, also zu Ausschuß führen. Der Ausschuß wird um so höher sein, je mehr Bauelemente im Schaltkreis integriert sind und je größer die Fläche des Chips ist, auf der der Schaltkreis untergebracht ist. Letzteres ist darauf zurückzuführen, daß die Anzahl der Chips ohne Material- und Bearbeitungsfehler abnimmt, wenn die Chipfläche wächst, vorausgesetzt, die Fehler sind auf der Substratscheibe gleichmäßig verteilt. Bei der Herstellung eines hochintegrierten Schaltkreises können in der Anfangsphase ohne weiteres 90° und mehr Ausschuß entstehen, aber ohne weiteres auch verkraftet werden. Bedeuten 10% Ausbeute doch, daß aus einer Siliziumscheibe einige zehn funktionierende Schaltkreise gewonnen werden. Wie in mancher anderen Beziehung gelten also auch in bezug auf den Ausschuß und die Produktionsausbeute in der Mikroelektronik völlig andere Maßstäbe als sonst in der Industrie. Wie überall in der Industrie können und müssen aber auch in der Mikroelek-

tronik Anstrengungen zur Erhöhung der Produktionsausbeute durch bessere Beherrschung der Technologie unternommen werden, nur daß es dabei nicht um eine Steigerung von 90 % auf 95 % geht, sondern um Steigerungen von 5 % auf 15 % bei sehr hochintegrierten Schaltkreisen oder von 40 % auf 80 % bei niedriger integrierten.

Die Gesamtheit der Prozeßschritte, denen eine Siliziumsubstratscheibe ausgesetzt werden muß, bis auf ihr die Schaltkreise erzeugt und geprüft sind, bezeichnet man, wie bereits erwähnt, mitunter als *Zyklus 1* der Halbleitertechnologie. Der Begriff «Zyklus» ist etwas irreführend, weil es dabei überhaupt nicht um einen Kreisprozeß geht, sondern um einen Prozeß, der in gerader Linie auf das Ziel, die Schaltkreise auf der Scheibe, fortschreitet. Dem Zyklus 1 vorgelagert ist die Herstellung der Siliziumsubstratscheiben, letztlich ausgehend von dem Rohmaterial Quarz, und die Anfertigung der Fotoschablonen für die Fotolithographie. letztlich ausgehend von den Funktionsanforderungen an den Schaltkreis. Der Weg von der allgemeinen Beschreibung der Funktion des Schaltkreises über die Herstellung eines logischen Schaltbildes, seine Umsetzung in eine Schaltung mit elektronischen Bauelementen bis hin zu den Vorlagen mit den Bauelementestrukturen, die für die Herstellung der Fotoschablonen benötigt werden, also die gesamte Entwicklungsarbeit für einen Schaltkreis, wird als *Schaltkreisentwurf* bezeichnet. Schaltkreisentwurf, Herstellung der Schablonen und Substratherstellung sind die drei wesentlichen Arbeitsgänge, die durchlaufen werden müssen, bevor

Tab. 8 Herstellungstechnologie der Mikroelektronik im Überblick

Zyklus 0	Ausgangsmaterialien Geforderte Schaltkreisfunktionen	Substratherstellung Schaltkreisentwurf Fotomaskenherstellung
		
Zyklus 1	Substratscheiben Fotomasken	Oxidation Fotolithographie Ätzen Dotierung Epitaxie Metallisierung Test
		
Zyklus 2	Scheibe mit Schaltkreisen	Vereinzeln Trägermontage Bonden Verkappen Test
		
	Schaltkreis	

der Zyklus 1 beginnen kann. Man faßt sie manchmal unter dem Begriff «Zyklus 0» zusammen. Alle Arbeiten, die an den Zyklus 1 angeschlossen werden müssen, um zum verkaufsfertigen Schaltkreis zu kommen, gehören zum Zyklus 2 der Halbleitertechnologie. Der Zyklus 2 beginnt mit dem Heraustrennen der Plättchen aus der Scheibe und endet mit dem Funktionstest am fertigen Schaltkreis. In Tabelle 8 ist der gesamte Herstellungsprozeß der Mikroelektronik noch einmal im Überblick dargestellt. Wir sind dem Leser einige weitere Erläuterungen zum Zyklus 0 und

zum Zyklus 2 schuldig. Wir werden uns dabei aber kürzer fassen als beim Zyklus 1, den wir als zentralen Teil der Herstellungstechnologie von integrierten Schaltkreisen mit voller Absicht ausführlicher behandelt haben. Wir beginnen mit dem *Zyklus 0*.

Zyklus 0:
Schaltkreisentwurf
 Der Schaltkreisentwurf hat die Bereitstellung der Vorlagen für die Fotoschablonen zum Ziel. Wir wollen uns diese Vorlagen zunächst als Zeichnungen auf Papier vorstellen, obwohl es, wie wir gleich noch sehen werden, gar nicht unbedingt Zeich-

nungen sein müssen. Für jede Schablone ist *eine* Zeichnung erforderlich. Auf jeder Zeichnung ist die horizontale Struktur einer bestimmten Materialschicht des Schaltkreises dargestellt, d. h., die vom betreffenden Material überdeckten geometrischen Bereiche sind nach Form und Lage aufgezeichnet. Der vollständige Satz dieser Zeichnungen für einen Schaltkreis wird als dessen «Layout» bezeichnet. Das Layout enthält alle Angaben über den vertikalen Aufbau des Schaltkreises aus den verschiedenen Materialschichten und über die geometrische Struktur jeder einzelnen Materialschicht. Der Ausgangspunkt für die Herstellung des Layout sind Forderungen über die Funktion des Schaltkreises, wie etwa, daß es ein Speicherschaltkreis für 64 KBit sein soll, daß jede Speicherzelle direkt adressierbar sein muß, daß die Betriebsspannung 5 Volt betragen soll, der Leistungsverbrauch nicht größer als 200 Milliwatt und die Zugriffszeit zu den gespeicherten Daten nicht größer als 100 Nanosekunden sein darf. Dazu kommen Forderungen allgemeiner Art, wie etwa die nach möglichst geringer Chipfläche, was u. a. bedeutet, daß die Zahl der äußeren Anschlüsse mit ihren platzaufwendigen Bondinseln möglichst klein gehalten werden muß. In Kapitel 4 haben wir auf den hohen Aufwand zum Test eines hochintegrierten Schaltkreises bereits hingewiesen. Diesen Aufwand in Grenzen zu halten, wird zunehmend zu einer Forderung schon beim Schaltkreisentwurf. Eine wichtige Vorgabe für den Entwurf ist die Entscheidung für eine der verschiedenen Realisierungstechniken. Zu jeder dieser Techniken gehören be-

stimmte Entwurfsregeln, die festlegen, welche minimalen Größen und Abstände und welche maximalen Überlappungen der einzelnen Strukturelemente nicht unter- bzw. überschritten werden dürfen. Aus der Gesamtheit dieser Regeln und Forderungen müssen die Zeichnungen für das Layout gewonnen werden. Am besten wäre es natürlich, wenn man diese Aufgabe voll und ganz einem Computer übertragen könnte. Dazu müßte ein Computerprogramm aufgestellt werden, mit dessen Hilfe die Koordinatenwerte der Punkte des Layout, das allen gestellten Anforderungen genügt, berechnet werden könnten. Prinzipiell ist das nichts Unmögliches. Praktisch ist man zur Zeit aber noch weit von einer solchen Lösung entfernt. Die Erfahrung und Intuition des Menschen, das Ausprobieren von verschiedenen Varianten und ihre schrittweise Verbesserung spielen noch eine wichtige Rolle beim Schaltkreisentwurf. Zur Lösung einiger aufwendiger Teilaufgaben kann aber ein Computer herangezogen werden. Man spricht in diesem Zusammenhang vom *computergestützten Entwurf* eines Schaltkreises.

In der Regel wird ein Schaltkreis stufenweise entworfen. Man beginnt auf der Logik-Stufe, d. h. mit einer Schaltskizze aus Symbolen für die elementaren logischen Operationen NAND, NOR, Negation usw. In Kapitel 4 haben wir Beispiele für logische Schaltungen kennengelernt. Der Test, ob die logische Schaltung wirklich die geforderten Funktionen erfüllt, die sogenannte Logik-Simulation, wird von einem Computer ausgeführt. Nach erfolgreichem Test wird die logische Schal-

tung mit Hilfe von elektronischen Bauelementen, Logik-Gattern oder größeren elektronischen Funktionsblöcken, als *elektronische* Schaltung auf dem Papier realisiert. Aus den elektrischen Kenngrößen der einzelnen Bauteile können die elektrischen Kenngrößen der Schaltung wie Spannungspegel, Leistungsaufnahme oder Verarbeitungsgeschwindigkeit berechnet werden. Auch diese Aufgabe, die Simulation der elektronischen Schaltung, die sogenannte Netzwerksimulation, wird vom Computer gelöst. Ist die Schaltung berechnet und stimmen ihre elektrischen Werte zunächst nicht mit den geforderten überein, muß man die Kenngrößen der elektronischen Bauteile ändern und die Rechnung mit den neuen Größen wiederholen. Unter Umständen muß aber auch die logische Schaltung verbessert werden. Hat man nach mehreren Versuchen eine geeignete Schaltung aus Bauteilen mit bestimmten Kenngrößen ermittelt, muß ein Layout für die Bauteile entworfen werden, das die Realisierung dieser Kenngrößen gewährleistet. Die Bauteile brauchen dazu nicht wirklich hergestellt zu werden. Ihre Kenngrößen lassen sich aus dem Layout ebenfalls mit Hilfe eines Computers berechnen. In vielen Fällen wird man aber auch auf bereits früher erprobte, in einer Bibliothek abgelegte Layouts für Bauelemente, Logik-Gatter oder größere Funktionsblöcke zurückgreifen können. Sind zufriedenstellende Layouts für die einzelnen Teile des Schaltkreises erreicht bzw. waren sie von Anfang an vorhanden, kann das Layout für den Gesamtschaltkreis in Angriff genommen werden. Für einen hochintegrierten Schalt-

kreis enthält es mehrere Millionen Punkte und Strecken. Für jede Zeichnung des Layout ist ein Papierbogen von mehreren Quadratmetern Fläche, also von der Größe eines mittleren Wandgemäldes, erforderlich. Die Fülle von Details auf den Zeichnungen übertrifft die in den Bildern von Peter Breughel um ein Vielfaches. Allerdings ist die Einförmigkeit der Zeichnungen sehr groß — es kommen im wesentlichen nur Rechtecke von unterschiedlicher Größe und Lage vor. Die Zeichnungen müssen auf etwa einen Zehntel Millimeter genau sein. Diese Forderung ergibt sich daraus, daß bei einem Schaltkreisplättchen von 4 mm Länge und 4 mm Breite die Toleranzen für die Bauelementestrukturen nicht mehr als einige Zehntel Mikrometer betragen dürfen. Nehmen wir $0,4\text{ }\mu\text{m}$ dafür an und setzen dazu die 4 mm der Gesamtabmessungen des Schaltkreises ins Verhältnis, so ergibt sich eine relative Zeichengenauigkeit von 10^{-4} . Bei einer Zeichnung von $1\text{ m} \times 1\text{ m}$ bedeutet das also eine Genauigkeit von 0,1 mm. Ein Beispiel für ein einfaches Layout ist in Abb. 49 dargestellt.

Das, was von den Zeichnungen auf Papier für die Herstellung der Fotoschablonen wirklich wesentlich ist, sind die Koordinaten der einzelnen Punkte und Strecken. Diese Koordinaten können nicht nur auf Papier in Form einer Zeichnung, sondern auch auf einem magnetischen Datenträger gespeichert werden. Davon wird heute weitgehend Gebrauch gemacht. Das heißt nicht, daß ein vorher auf Papier gezeichnetes Layout nachträglich auf ein Magnetband übertragen wird. Vielmehr wird bei

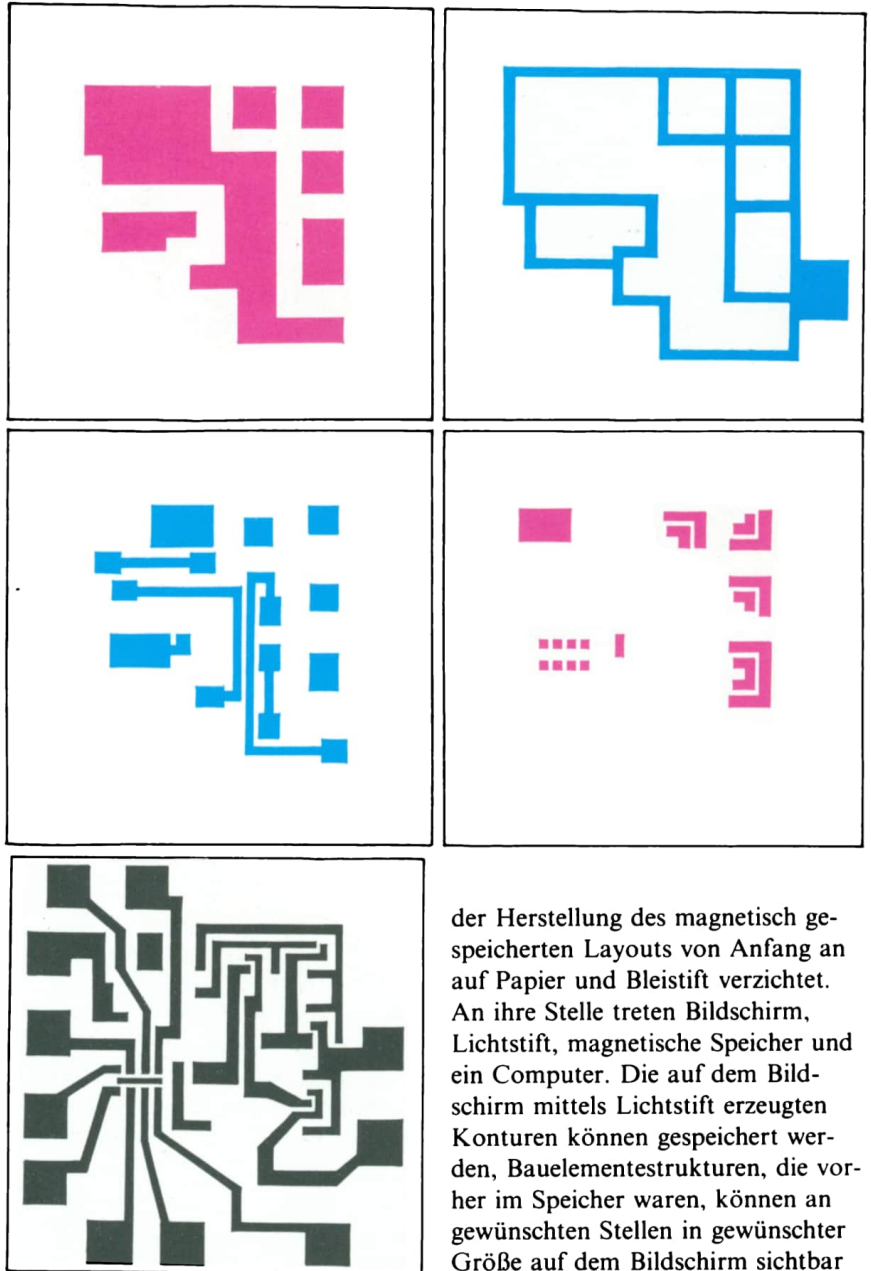


Abb. 49 Layout für einen TTL-Schaltkreis. Die roten Flächen entsprechen den n-Gebieten, die blauen den p-Gebieten und die schwarzen den leitenden Aluminium-Gebieten. Die Zeichnung für die Kontaktfenster ist nicht dargestellt.

der Herstellung des magnetisch gespeicherten Layouts von Anfang an auf Papier und Bleistift verzichtet. An ihre Stelle treten Bildschirm, Lichtstift, magnetische Speicher und ein Computer. Die auf dem Bildschirm mittels Lichtstift erzeugten Konturen können gespeichert werden, Bauelementestrukturen, die vorher im Speicher waren, können an gewünschten Stellen in gewünschter Größe auf dem Bildschirm sichtbar gemacht, verändert und untereinander kombiniert werden. Das magnetisch gespeicherte Layout kann unter Nutzung mikroelektronischer Informationstechnik in einer Weise

hergestellt werden, die der früher üblichen Zeichentechnik weit überlegen ist. Die Mikroelektronik ist auf diese Weise selbst zu einem wichtigen Instrument geworden, um neue mikroelektronische Schaltkreise zu entwickeln.

Nachdem ein befriedigendes Layout entworfen und auf dem Magnetband gespeichert ist, werden mit seiner Hilfe die Fotoschablonen hergestellt. Das Verfahren dafür werden wir weiter unten kurz besprechen. Hier nehmen wir an, daß die Schablonen bereits fertig sind. Sie werden benutzt, um die ersten Exemplare des Schaltkreises auf einer Siliziumscheibe zu erzeugen. Der elektronische Test der Schaltkreise gibt endgültig Auskunft über Erfolg oder Mißerfolg des Schaltkreisentwurfs. Erfüllt der Schaltkreis die Anforderungen nicht, beginnt die Arbeit von vorn. Meist sind bis zu fünf, unter Umständen aber auch mehr Wiederholungen der gesamten oben beschriebenen Prozedur notwendig, um zu einem Schaltkreis zu kommen, der den gestellten Anforderungen genügt. Nach diesen Ausführungen über den Schaltkreisentwurf wird es niemanden mehr verwundern, wenn wir feststellen, daß es sich dabei um einen sehr arbeitsintensiven Vorgang handelt. Es ist üblich, den Arbeitsaufwand in «Mannjahren» zu messen. Für einen Schaltkreis aus einigen Tausend Bauelementen mußte man vor noch nicht langer Zeit etwa ein Mannjahr Entwurfsarbeit veranschlagen, d. h., ein Mann mit der erforderlichen Qualifikation mußte ein Jahr lang am Entwurf des Schaltkreises arbeiten. Für einen Mikroprozessor mit drei Schaltkreisen aus

je 100 000 Bauelementen wurden 100 Mannjahre Entwurfsarbeit angegeben. Dieser hohe Aufwand wird sich in dem Maße weiter reduzieren, in dem weitere Teilaufgaben an den Elektronen-Rechner übertragen werden können.

Fotoschablonen

Wir haben eine Erläuterung zur Herstellung der Fotoschablonen nachzutragen. Ausgangspunkt dafür waren früher Lack-beschichtete Folien von etwa 1 m² Fläche. Die Lackschicht war abziehbar und undurchsichtig, die Folie durchsichtig. Mittels eines durch die Layout-Daten gesteuerten Koordinatenschreibers, der statt mit einem Stift mit einem feinen Messer ausgerüstet war, wurden die Layout-Konturen in die Lackschicht eingeritzt. Anschließend konnten die ausgeschnittenen Lackflächen mit einer Pinzette abgehoben werden, wodurch gewissermaßen Fenster in der Folie entstanden. Durch Verkleinerung und Übertragung der so erzeugten Fensteranordnung auf eine Fotoplatte entstanden die sog. Zwischenschablonen (englisch Reticles) für die verschiedenen Materialebenen eines Chips. Heute benutzt man anstelle dieser Technik meist ein Licht- oder Elektronenoptisches Verfahren. Das Reticle wird nicht auf dem Umweg über eine Lack-beschichtete Folie erzeugt, sondern durch direkte Belichtung einer Fotoplatte, die oft noch mit Chromoxid überzogen ist (Chromplatte). An die Stelle des Koordinatenschreibers tritt ein Licht- oder Elektronenoptischer Pattern(Muster)generator. Nur auf diesem Wege lassen sich die Reticles für hochintegrierte Schaltkreise rationell herstellen. Die Reticles werden verwendet, um die sog. Originalschablonen zu erzeugen. Das

sind Foto- oder Chromplatten, die in Originalgröße alles enthalten, was später bei der Fotolithografie auf der Substratscheibe abgebildet werden soll, also nicht nur die Strukturen für *einen* Schaltkreis, sondern für *so viele*, wie auf der Scheibe Platz haben, und darüber hinaus noch Positioniermarken. Die Vervielfältigung und Verkleinerung der Reticles erfolgt mit dem sog. Photorepeater.

Für die Fotolithografie werden nicht die Originalschablonen selbst verwendet, sondern sog. Arbeitsschablonen, die man durch Kopieren aus den Originalschablonen gewinnt. Bei der Belichtung des Fotolacks auf der Siliziumscheibe legt man die Arbeitsschablonen entweder direkt auf die Scheibe auf oder man projiziert sie, meist im Maßstab 1:1. Beim Projektionsverfahren werden die Schablonen wesentlich schonender behandelt als beim direkten Auflegen, wo sie bereits nach zehn- bis zwanzigmaliger Verwendung abgenutzt sind. Anstelle von Licht können zur Belichtung bei der Lithografie auch Röntgen- oder Elektronenstrahlen eingesetzt werden. Mit Hilfe eines Elektronenstrahls lassen sich auch ohne die Verwendung von Schablonen Strukturen erzeugen. Man bezeichnet dieses Verfahren als Raster-Elektronenstrahlolithografie. Dabei werden die Koordinatenwerte des Layout zur Positionierung eines feinen Elektronenstrahls auf der Lack-beschichteten Siliziumscheibe verwendet. Der Elektronenstrahl schreibt das Layout auf die Siliziumscheibe auf. Dieses Verfahren ist aber sehr zeitaufwendig und wird nur in speziellen Fällen angewandt — im Unterschied zur Schablonenherstellung, wo Elektronenstrahlen heute breite Verwendung finden.

Bei den Ausführungen zum Schaltkreisentwurf und zur Schablonenherstellung haben wir so getan, als ob die Substratscheiben bereits vorhanden wären. Wie wir dazu kommen, soll jetzt besprochen werden.

Substratscheiben

Für die Herstellung der Substratscheiben benötigt man chemisch reines Silizium als Ausgangsmaterial. Es wird aus Quarz, d. h. kristallisiertem Siliziumdioxid, durch Reduktion mit Kohlenstoff gewonnen. Dabei verbindet sich der Sauerstoff des Siliziumdioxids mit dem Kohlenstoff zu Kohlenmonoxid und das Silizium bleibt als reines Element übrig. Die beiden Ausgangsmaterialien für diese chemische Reaktion, Quarz und Kohlenstoff, werden von der Natur in ausreichender Menge zur Verfügung gestellt. Rohstoffprobleme gibt es also für die Mikroelektronik insoweit nicht, wie es dabei nur um die verfügbaren Mengen geht. Silizium ist nach Sauerstoff das zweithäufigste Element der Erdkruste, die zu 28 % aus Silizium besteht. Die Probleme mit den Ausgangsmaterialien liegen auf einer anderen Ebene. Es geht um ihre Reinheit, die nur um einen hohen Preis zu erreichen ist. Immerhin entsteht durch die Reduktion des Quarzes mit Hilfe von Kohlenstoff ein Silizium, das bereits einen Reinheitsgrad von 98–99 % besitzt. Bevor man aber Kristalle daraus züchten kann, muß das Silizium noch weiter gereinigt werden. Die Beseitigung der störenden Schwermetallatome wie Kupfer oder Eisen geschieht auf folgende Weise: Mit Salzsäure wird das Silizium in Trichlorsilan (SiHCl_3) umgewandelt. Die Schwermetallatome ver-

binden sich dabei mit dem Chlor zu Chloriden. Bei der anschließenden fraktionierten Destillation verdampfen die Chloride. Das verbleibende Trichlorsilan wird mit Wasserstoff gemischt und bei etwa 1000 °C reduziert. Das Silizium selbst scheidet sich auf einem dünnen Siliziumstab ab, den man als Kernmaterial vorher in den chemischen Reaktor — ein langes Quarzrohr — einbringen muß. Das Ergebnis sind Siliziumstäbe von etwa einem Meter Länge und acht Zentimetern Dicke. Es schließen sich physikalische Reinigungsverfahren an. Dabei wird ausgenutzt, daß die Aufnahmefähigkeit des Siliziums für Fremdstoffe in der flüssigen Phase größer ist als in der festen. Erzeugt man also in einem Siliziumstab — etwa durch Aufheizung mit einer Induktionsspule über 1420 °C, dem Schmelzpunkt des Siliziums — eine flüssige Zone, so nimmt diese die Verunreinigungen aus den angrenzenden festen Bereichen des Stabes in sich auf. Wird die Zone langsam über den Stab hinweggezogen, indem die Induktionsspule bewegt wird, so führt sie die Verunreinigungen mit sich. Am Stabende werden sie abgelagert. Führt man diesen Prozeß mehrfach durch, so entsteht ein hochreiner Siliziumstab, in dem, bis auf das schmutzige Stabende, die Konzentration an kritischen Verunreinigungen auf 10^{-9} , das entspricht 0.1 millionstel Prozent, herabgedrückt ist. Dieses Reinigungsverfahren bezeichnet man als *Zonenschmelzen* (siehe Abb. 50, S. III). Daran schließt sich die eigentliche Kristallherstellung an. Ihr Ziel besteht darin, aus dem zwar reinen, vom atomaren Aufbau her aber noch ungeordneten Siliziumstab einen Kristall zu machen. Bei

einem Kristall sind die Atome absolut regelmäßig angeordnet, beim Silizium auf den Ecken eines regelmäßigen Tetraeders, in dessen Mittelpunkt sich ebenfalls ein Siliziumatom befindet (siehe Abb. 8). Die Gefahr, daß bei der Kristallzüchtung Abweichungen von diesem idealen Aufbau, z. B. Versetzungen entstehen, ist relativ groß. Bei der Herstellung der Siliziumkristallstäbe für die Mikroelektronik muß man sie soweit wie irgend möglich vermeiden. Dazu dienen zwei Kristallzüchtungsverfahren, das Czochralski-Verfahren und das Zonenschmelz-Verfahren (Abb. 50). Bei beiden Verfahren arbeitet man mit einem Kristallkeim, einem perfekt gebauten kleinen Kristall, den man beim Czochralski-Verfahren in eine Siliziumschmelze eintaucht und danach zusammen mit dem anhängenden Material unter ständiger Rotation langsam wieder herauszieht. Bei der sich anschließenden Abkühlung kristallisiert dieses Material in einer Weise, die durch den Kristallkeim bestimmt wird. War der Keim perfekt, wird der aus ihm wachsende Kristallstab auch perfekt sein — die Einhaltung der optimalen Bedingungen für den Kristallisationsprozeß vorausgesetzt. Die Ziehgeschwindigkeiten beim Czochralski-Verfahren liegen im Bereich von einigen Zehntel Millimetern bis zu einigen Millimetern pro Minute. Beim Zonenschmelz-Verfahren wird der Keim am oberen Ende eines Siliziumstabes, wie er nach Abschluß der Reinigungsverfahren und nach Abtrennung der verunreinigten Enden vorliegt, befestigt. Von oben her wird eine Schmelzzone langsam über den Stab hinweggezogen. Hinter ihr erstarrt

das Silizium zu einem Einkristall. Die Oberflächenspannung hindert das aufgeschmolzene Silizium daran, zur Seite wegzufließen, allerdings nur so lange, wie der Siliziumstab nicht zu dick ist. Mit dem Czochralski-Verfahren können heute Kristallstäbe mit einem Durchmesser bis zu 150 mm ohne Veretzungen gezüchtet werden. Mit dem Zonenschmelz-Verfahren erreicht man Durchmesser bis zu 100 mm, ebenfalls ohne Versetzungen. Gemessen an natürlich gewachsenen Kristallen ist das ein sehr gutes Ergebnis. Völlig vermeiden kann man Defekte aber auch bei diesen Verfahren nicht — es entstehen Punktdefekte, z. B. Vakanzen. Auch chemische Verunreinigungen sind nicht völlig zu unterdrücken. Das nach dem Czochralski-Verfahren hergestellte Material hat im Unterschied zum Material aus dem Zonenschmelz-Verfahren sogar einen relativ hohen Gehalt an Sauerstoff- und Kohlenstoffatomen (bis zu 10^{18} cm^{-3}). Diese Atome stammen aus dem beim Czochralski-Verfahren unumgänglichen Quarztiegel zur Aufbewahrung der Siliziumschmelze und dem zugehörigen Graphit-Heizer. Mit Luft kommen die Stäbe natürlich während des gesamten Herstellungsvorganges nicht in Berührung. Der Prozeß findet in einer schützenden Edelgasatmosphäre statt. Beim Zonenschmelz-Verfahren wird völlig ohne Tiegel gearbeitet, der Gehalt an Sauerstoff und Kohlenstoff ist deshalb geringer. Beide Elemente können aber im Unterschied zu Schwermetallen selbst in einer Konzentration von 10^{18} cm^{-3} von der Mikroelektronik toleriert werden. Der Sauerstoff hat sogar noch eine positive Wirkung: Er kann ausgenutzt

werden, um die störenden Schwermetallatome unwirksam zu machen. Diesen Effekt bezeichnet man als (innere) *Getterung*.

Die Einkristallstäbe werden in 0.5 bis 1 mm dicke Scheiben zersägt. Die dazu verwendeten Sägeblätter (z. B. mit Diamantenstaub belegte Metallscheiben) sind fast genauso dick wie die Scheiben selbst. Aus den kostbaren Siliziumstäben werden also fast zur Hälfte „Sägespäne“. Die Oberflächenschichten der Scheiben, auf die es beim Zyklus 1 vor allem ankommt, sind durch das Sägen stark gestört worden. Vor der Verarbeitung im Zyklus 1 müssen daher die gestörten Oberflächenbereiche zunächst entfernt werden. Das geschieht durch Schleifen, Läppen, Ätzen und Polieren. Das Ergebnis sind schwarz glänzende Substratscheiben, wie sie z. B. in Abb. 42 zu sehen sind. Sie bilden den Abschluß des Zyklus 0 und den Ausgangspunkt für den Zyklus 1, den wir bereits ausführlich behandelt haben.

Es bleibt uns jetzt noch, auf den Zyklus 2 der Halbleitertechnologie etwas genauer einzugehen. Er führt von den Scheiben mit den getesteten Schaltkreisstrukturen zu den ver-

Zyklus 2:

Vereinzeln der Plättchen
kaufsfertigen Schaltkreisen. Der erste Prozeßschritt des Zyklus 2 ist das Zertrennen der Scheibe in die einzelnen Schaltkreislplättchen oder Chips. Dazu wird die Scheibe mit einem Diamanten entlang der späteren Plättchenkanten geritzt und anschließend auf eine Polyäthylenfolie geklebt. Durch Biegen und Dehnen der Folie zerreißt die Scheibe längs der Ritze, so wie das in Abb. 51 (siehe S. VI) zu sehen ist. Von

jetzt an muß jedes Plättchen einzeln weiter verarbeitet werden. Vorher, im Zyklus 1, waren es Scheiben mit hunderten von Schaltkreisen, die gleichzeitig einem Prozeßschritt unterworfen wurden. Jetzt, im Zyklus 2, muß jeder Arbeitsvorgang, je nach der Höhe der Ausbeute, zehn bis hundertmal häufiger ausgeführt werden als im Zyklus 1. Die Automatisierung der Arbeiten im Zyklus 2 ist deshalb von ganz besonderer Bedeutung.

Bonden

Nach dem Vereinzeln werden die funktionstüchtigen Chips auf einem stabilen Träger befestigt und in ein Gehäuse eingehüllt, an dem auch die äußeren Anschlußpole des Bausteins angebracht sind. Bei Gehäusen aus Keramik ist der Träger meist der Keramikboden des Gehäuses. In zunehmendem Maße wird heute aber kein Gehäuse im eigentlichen Sinne mehr verwendet, sondern der Chip wird in eine Plastmasse eingegossen. In diesem Falle befestigt man den Chip im Zentrum eines metallischen Trägers (*Chipbonden*), so wie es in Abb. 52 (siehe S. VII) zu sehen ist. Die auf das Plättchen zeigenden vergoldeten Finger in Abb. 52 gehen nach außen in die Anschlußkontakte des Bausteins über. Man hat in Abb. 52 gleich mehrere solche Träger miteinander zu einem sog. *Trägerstreifen* verbunden. Auf das Chipbonden folgt das *Drahtbonden*, d. h. das leitende Verbinden der Bondinseln auf dem Schaltkreislplättchen mit den Anschlußfingern des Trägers bzw. mit den Anschlußpolen des Keramik-Gehäuses, falls ein solches verwendet wird. Dazu werden feine Golddrähtchen mittels unterschiedlicher Verfahren auf die

Kontaktstellen aufgeschweißt. Diese Tätigkeit wurde früher manuell unter einem Stereomikroskop ausgeführt. Heute werden weitgehend Automaten dafür eingesetzt. Das Ergebnis, ein Kontakt zwischen dem Golddrähtchen und der Bondinsel aus Aluminium, ist in Abb. 53 (siehe S. VII) dargestellt. Beim anschließen-
Verkappen

den Verkappen bleiben die Verbindungsteile zwischen den einzelnen Trägern weiter bestehen. Es folgt eine elektrische Kontrolle der gebondeten Chips und, bei positivem Ergebnis, das Umhüllen mit einem Plastmaterial. Man verwendet dazu vorwiegend Epoxidharze. Bei etwa 300 °C werden die Träger mit den Chips in das Harz eingegossen. Die Anschlußpole müssen dabei natürlich frei bleiben. Danach entfernt man alle nicht mehr benötigten Teile des Trägerstreifens einschließlich der Verbindungen zwischen den einzelnen Trägern. Die Gehäuseanschlüsse werden nach unten gebogen, und der Schaltkreis im Plastgehäuse, so wie ihn Abb. 54 (siehe S. XI) in einigen Exemplaren zeigt, ist fertig. Er muß nun einer gründlichen Prüfung unterzogen werden, und zwar nicht nur im Hinblick auf sein einwandfreies elektrisches Funktionieren, sondern auch im Hinblick auf seine Temperatur- und mechanische Stabilität. Erst wenn der elektrische Test auch noch nach erfolgter Wärme- und mechanischer Belastung positiv ausfällt, ist der Schaltkreis verkaufsfertig.

Herstellungskosten

Vor dem Verkauf müssen seine Herstellungskosten festgestellt werden. Auf einige Gesichtspunkte, die dabei eine Rolle spielen, soll zum Abschluß dieses Kapitels kurz eingegan-

gen werden. Die Gesamtkosten für die Herstellung eines Schaltkreises sind wesentlich durch die Kosten bestimmt, die in den Zyklen 1 und 2 entstehen. Die Kosten für die Herstellung der Substratscheiben im Zyklus 0 gehen in die Aufwendungen für den Zyklus 1 mit ein. Der Schaltkreisentwurf und die Schablonenherstellung wirken sich auf die Gesamtkosten um so weniger aus, in je größerer Stückzahl der Schaltkreis gefertigt wird. Die in den Zyklen 1 und 2 anfallenden Kosten hängen ihrerseits stark vom Integrationsgrad ab. Daß sie bei einem höher integrierten Schaltkreis größer sein werden als bei einem niedriger integrierten liegt auf der Hand: Die Ausbeute an funktionierenden Schaltkreisen wird geringer sein, die Fläche des Chips und damit die Kosten für das Siliziumsubstrat werden höher liegen und der Aufwand für Funktionstests wird umfangreicher sein, wenn der Integrationsgrad zunimmt. Dafür ist aber auch die Leistungsfähigkeit des höher integrierten Schaltkreises größer als die des niedriger integrierten. Eine richtige ökonomische Bewertung liefern überhaupt nicht die absoluten Kosten eines Schaltkreises, sondern die auf seine Leistungsfähigkeit bezogenen. Die Leistungsfähigkeit eines Schaltkreises ist vor allem durch seinen Integrationsgrad, d. h. durch die Anzahl der in ihm enthaltenen Bauelemente, bestimmt. Ökonomisch interessant sind deshalb die Kosten des Schaltkreises dividiert durch die Zahl der in ihm enthaltenen Bauelemente, also die Kosten für ein integriertes Bauelement.

Für ökonomische Bewertungen in der Mikroelektronik ist es sehr wich-

tig, daß und wie diese Kenngröße «Kosten eines integrierten Bauelements» vom Integrationsgrad abhängt (Abb. 55). Sie durchläuft ein

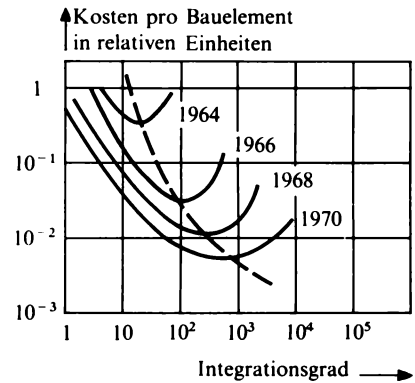


Abb. 55 Abhängigkeit der Herstellungskosten bezogen auf ein integriertes Bauelement bei einem Schaltkreis in Abhängigkeit vom Integrationsgrad

ausgeprägtes Minimum, d. h., bei einem bestimmten Integrationsgrad sind die Kosten eines integrierten Bauelementes minimal. Wird der optimale Integrationsgrad über- oder unterschritten, steigen die Kosten an. Liegt der Integrationsgrad unterhalb des optimalen Wertes, führt eine Erhöhung der Integration zu einer Kostensenkung. Befindet man sich genau beim optimalen Wert und erhöht die Ausbeute, was durch den «Lernprozeß» bei der Herstellung bis zu einem gewissen Grad automatisch geschieht, so tritt das Kostenminimum bei einem höheren Integrationsgrad auf und die Minimalkosten selbst werden geringer. Will man also den Kostenvorteil einer erhöhten Ausbeute voll ausschöpfen, muß man gleichzeitig höher integrieren. Daraus erklärt sich zu einem

guten Teil die Tendenz der Mikroelektronik, zu Schaltkreisen mit immer höheren Integrationsgraden überzugehen. Seit den Anfängen der Mikroelektronik im Jahre 1960 wurde der Integrationsgrad jährlich etwa verdoppelt. Welche geradezu spektakulären Kostensenkungen damit erreicht werden konnten, zeigt die Entwicklung bei den Speicherschaltkreisen, wo die Kosten eines integrierten Bauelements 1980 tausendmal niedriger waren als 1970. Hätte die Autoindustrie ihre Produkte in gleichem Maße verbilligen können, so würde heute ein Auto etwa 10 M kosten. Allerdings wäre es auch nur noch so groß wie eine Streichholzschachtel, wenn die Verbilligung auf dem gleichen Wege herbeigeführt worden wäre wie sie in der Mikroelektronik erzielt wurde. Man erkennt auch daraus wieder, daß dem Vergleich der Mikroelektronik mit anderen Industriezweigen Grenzen gesetzt sind. Die Herstellungstechnik der Mikroelektronik ist in vielfacher Hinsicht einmalig und unvergleichbar. Neben vielem Einzelnen war das die allgemeine Aussage des vorliegenden Kapitels.

6. Integrierte Schaltkreise und ihre Anwendung

Wie Halbleiterbauelemente funktionieren, wie sich mit ihnen Schaltungen für die Informationsverarbeitung aufbauen lassen und wie integrierte Schaltkreise hergestellt werden können, haben wir in den vorangegangenen Kapiteln kennengelernt. In diesem Kapitel soll besprochen werden, welche Arten von integrierten Schaltkreisen die Industrie heute tatsächlich produziert und warum diese und keine anderen im Angebot sind. Wir orientieren uns dabei an den digitalen Schaltkreisen, ohne die analogen ganz auszuklammern.

Ein passender Schaltkreis für jedes Problem?

Jedes Informationsverarbeitungsproblem läßt sich auf die Ausführung elementarer logischer Operationen mit 0- und 1-Impulsen zurückführen, die zur Lösung des Problems in bestimmter Weise miteinander verknüpft werden müssen. In der Art dieser Verknüpfung besteht die spezielle Natur des Problems. Wie wir gesehen haben, können die elementaren logischen Operationen durch einfache Schaltungen realisiert werden, und die Verknüpfung der logischen Operationen bedeutet nichts weiter, als diese einfachen Schaltungen untereinander zu einer einzigen großen Schaltung zu verbinden. Was wir im vorangegangenen Kapitel am Beispiel einiger spezieller Informationsverarbeitungs-

probleme (z. B. Zählen von Impulsen, Addition oder Speicherung von Binärzahlen) kennengelernt haben, gilt ganz allgemein: Für jedes Informationsverarbeitungsproblem gibt es eine geeignete Schaltung, die das Problem zu lösen imstande ist. Würde diese dem Problem auf den Leib geschriebene Schaltung von der Mikroelektronik-Industrie tatsächlich als integrierter Schaltkreis angeboten werden, so wäre das für den Anwender sehr bequem. Die Schaltung würde ihm in einem Gehäuse mit den erforderlichen Anschlüssen zur Verfügung stehen. Er brauchte diese Anschlüsse nur in der richtigen Weise mit den Leitungen für die Zuführung der zu verarbeitenden Information, für die Abnahme der verarbeiteten Information und für die elektrische Versorgungsspannung zu verbinden, und das System zur Lösung seines Informationsverarbeitungsproblems wäre fertig. Die Situation wäre ähnlich wie beim Kauf eines Autos. So wie das Auto von der Industrie hergestellt wird, befriedigt es das Bedürfnis des Kunden nach einem Fortbewegungsmittel vollständig. Der Kunde muß nicht erst Teile zusammenbauen oder das, was die Industrie liefert, seinem Bedürfnis nach Fortbewegung anpassen. Für die Mikroelektronik-Industrie ist dieser Weg im allgemeinen aber weder gangbar noch

effektiv. Dafür sind vor allem drei Gründe ausschlaggebend.

Problemkomplexität und Integrationsgrad

Erstens sind viele praktisch wichtige Informationsverarbeitungsprobleme viel zu kompliziert, d. h., ihre Lösung erfordert viel zu viele aufeinanderfolgende elementare logische Verknüpfungsoperationen, um mit einem einzigen integrierten Schaltkreis bewältigt werden zu können. Für einen großen elektronischen Rechner z. B. benötigt man die Funktionen von . . . zig Millionen Bauelementen. Aber auch die höchstintegrierten Schaltkreise enthalten heute noch weniger als eine Million integrierte Bauelemente. Die Schaltung für einen großen Rechner läßt sich weder heute noch in absehbarer Zukunft auf einem einzigen Halbleiterchip unterbringen. Sie kann nur dadurch realisiert werden, daß eine große Zahl integrierter Schaltkreise zusammengeschaltet wird. Was über die Schaltung für einen elektronischen Rechner gesagt wurde, gilt auch für Schaltungen zur Lösung vieler anderer Informationsprobleme; auch sie lassen sich beim gegenwärtigen internationalen Stand der Technik nur durch nachträgliches Zusammenschalten einzelner integrierter Schaltkreise aufbauen. Das Zusammenschalten wird, wenn es sich, wie beim elektronischen Rech-

ner, um das Informationsverarbeitungsproblem eines kommerziell gefertigten Gerätes handelt, von einem industriellen Hersteller besorgt. Schaltungen zur Lösung kleinerer Informationsprobleme können aber auch von einem vorgebildeten Nutzer selbst aus den von der Industrie angebotenen integrierten Schaltkreisen aufgebaut werden. Was sich früher, bevor es die Mikroelektronik gab, beim Aufbau einer Schaltung für ein bestimmtes Informationsverarbeitungsproblem auf der Ebene von Einzelbauelementen vollzog, geschieht heute auf der Ebene von integrierten Schaltkreisen. Der Fortschritt besteht in der Leistungsfähigkeit der Schaltungen, die natürlich unvergleichlich höher ist, wenn integrierte Schaltkreise statt der gleichen Anzahl einzelner Bauelemente zusammengeschaltet werden. Trotzdem werden hieran auch die Grenzen des heutigen internationalen Standes der Technik in der Mikroelektronik sichtbar. Es wäre besser, weil billiger, schneller und zuverlässiger, wenn die zur Lösung des Informationsverarbeitungsproblems eigentlich benötigte Schaltung nicht erst aus einzelnen kleineren integrierten Schaltkreisen nachträglich aufgebaut werden müßte, sondern von der Industrie schon als höher integrierter Schaltkreis geliefert werden könnte. Dabei geht es nicht so sehr um die *prinzipielle* Herstellbarkeit des integrierten Schaltkreises, mit hohem Aufwand läßt sich hier manches erreichen. In erster Linie kommt es auf seine Herstellungskosten, seine Leistungsparameter und seine Zuverlässigkeit an. Damit die Industrie einen neuen integrierten Schaltkreis auf den Markt

bringen kann, müssen in diesen Punkten gewisse Mindestanforderungen befriedigt werden. Erst dann kann man wirklich sagen, daß die Industrie technisch in der Lage ist, den Schaltkreis herzustellen.

Problemvielfalt und Programmierung

Aber selbst dann, wenn die Mikroelektronik-Industrie in diesem Sinne technisch dazu imstande wäre — wir kommen damit zum zweiten Punkt — wäre es praktisch unmöglich, für jedes Informationsproblem einen geeigneten integrierten Schaltkreis zur Verfügung zu stellen. Dafür ist die Anzahl der verschiedenen Informationsprobleme viel zu groß. Man muß sich dazu vergegenwärtigen, daß zwei Informationsverarbeitungsprobleme bereits dann nicht mehr identisch sind, wenn nur an einer einzigen Stelle in der Aufeinanderfolge der elementaren logischen Operationen, in die jedes Problem zerlegt werden kann, ein Unterschied besteht. Die Addition und Subtraktion zweier Zahlen sind in diesem Sinne bereits unterschiedliche Probleme. Die Kapazität der heutigen Mikroelektronik-Industrie wäre absolut überfordert, sollte sie so viele verschiedene Schaltkreise herstellen, wie es verschiedene Informationsprobleme gibt. Es wäre aber auch unmöglich, die Kapazitäten so zu erweitern, daß die Verfügbarkeit aller dieser Schaltkreise gewährleistet werden könnte. Dazu müßte ein viel zu großer Teil der Menschheit in der Mikroelektronik-Industrie arbeiten. Was ist in dieser Lage zu tun? Diese Frage führt uns auf den Begriff der *Programmierung* eines Informationsverarbeitungssystems. Vergegenwärtigen wir uns noch ein-

mal die Situation: Bestimmte Informationsverarbeitungsprobleme, wie oben die Addition und Subtraktion, unterscheiden sich nur an wenigen Stellen in der Aufeinanderfolge der elementaren logischen Operationen, während sie über weite Strecken übereinstimmen. Das bedeutet, daß sich auch die Schaltungen zur Lösung der beiden Informationsverarbeitungsprobleme weitgehend decken und nur an wenigen Stellen voneinander abweichen werden. Sollte es nicht möglich sein, diese weitgehende Übereinstimmung auszunutzen? Sollte man nicht eine einzige Schaltung aufbauen können, mit der sowohl die Addition als auch die Subtraktion durchgeführt werden kann, die aber trotzdem nur wenig komplizierter ist als die einzelne Schaltung für die Addition oder Subtraktion? Daß sie wenigstens etwas komplizierter sein müßte, ist klar. Es muß nämlich die Möglichkeit bestehen, der Schaltung zu befehlen, daß sie addieren soll, wenn die Addition gewünscht ist, und daß sie subtrahieren soll, wenn die Subtraktion gewünscht ist. Die Schaltung muß, wie man sagt, *programmierbar* sein. Das Prinzip der Programmierung ist für die Informationsverarbeitung so wichtig, daß wir uns dabei noch einen Moment aufhalten wollen. Es soll versucht werden, das Wesentliche dieses Prinzips anhand eines gegenständlichen Beispiels zu veranschaulichen. Bei Informationsverarbeitungsproblemen muß eine vorgegebene Eingangsinformation nach einem bestimmten Verarbeitungsschema in eine gewünschte Ausgangsinformation umgewandelt werden. Etwas Ähnliches ist beim Transport von

Waren zu leisten: Eine Ware ist von einem bestimmten Ausgangsort A zu einem Bestimmungsort B zu bewegen. Nehmen wir an, daß es in der Umgebung des Ortes A bestimmte Orte B, C, D usw. gibt, zu denen Waren von A zu transportieren sind (Abb. 56). Der Transport soll per Eisenbahn erfolgen. Schienen sind noch nicht vorhanden, sondern müssen erst gelegt werden. Wie wird man das Schienennetz konzipieren? Eine Möglichkeit wäre, von A aus zu allen Orten B, C, D usw. einen eigenen Schienenstrang zu legen (Bild a). Sind die Orte B, C, D nicht weit voneinander entfernt, so wird das sicher nicht die günstigste Lösung für das Schienennetz sein. Man kann Gleise dadurch einsparen, daß man bis zu einem bestimmten Punkt X dicht bei den Orten B, C, D nur einen einzigen Strang legt und erst von da an drei getrennte Gleise (Bild b) benutzt. Allerdings muß man dann im Punkt X eine verstellbare Weiche einbauen, um die Züge in die jeweils gewünschte Richtung lenken zu können. Durch die Weiche wird das Gleisnetz gewissermaßen programmierbar. Ob ein in A losgeschickter Zug in B, C oder D ankommt, kann durch die Stellung der Weiche vorprogrammiert werden. Die Beziehung zum programmierbaren Informationsverarbeitungssystem liegt auf der Hand. Die Züge entsprechen den Informationsströmen und die Schienen der Schaltung des Informationsverarbeitungssystems. Wie die Informations-Züge die Schaltungs-Schienen durchlaufen, wird durch die Programm-Weichen bestimmt. Ebenso wie es durch die verstellbaren Weichen möglich ist, ein und dasselbe Schie-

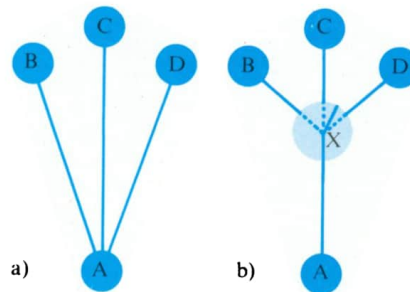


Abb. 56 Prinzip der Programmierung

nennetz für verschiedene Transportaufgaben zu nutzen, wird es durch die Programmierbarkeit möglich, ein und dasselbe Informationsverarbeitungssystem für verschiedene Informationsprobleme einzusetzen. Ebenso wie durch die Weichen bestimmte Schienenstränge mehrfach genutzt und Schienen eingespart werden können, lassen sich in einem programmierbaren Informationsverarbeitungssystem bestimmte Teilschaltungen vielfältig verwenden und ihre Zahl kann insgesamt kleiner gehalten werden als in einem nicht-programmierbaren System gleicher Leistungsfähigkeit. Die Programmierbarkeit ermöglicht es, die gleiche Informationsverarbeitungsleistung mit einem geringeren Aufwand an Hardware zu bewältigen. Hardware, d. h. die Gesamtheit der integrierten Schaltkreise, wird teilweise durch Software, d. h. durch ein Programm, ersetzt.

Auf einen weiteren Aspekt, in dem die Analogie zwischen Schienennetz und Informationsverarbeitungssystem zu richtigen Schlußfolgerungen führt, sei noch hingewiesen. Der Weg von A nach B über die Weiche bei X ist etwas länger als der direkte Weg von A nach B ohne Weichen. Auch bei der Lösung eines Informa-

tionsverarbeitungsproblems über ein programmierbares Verarbeitungssystem verliert man etwas von der Schnelligkeit, wie sie bei einem speziellen, nichtprogrammierbaren System möglich wäre. Das Prinzip der Programmierung von Informationsverarbeitungssystemen ist übrigens älter als man vermuten würde. Es wurde bereits 1833 von Charles Babbage beim Entwurf einer mechanischen Rechenmaschine, die allerdings nie gebaut wurde, verwendet. In seiner heutigen Form der Speicherprogrammierung, in der es in elektronischen Rechnern realisiert ist, wurde es erstmalig 1946 von J. von Neumann eingesetzt. Unser kurzer geschichtlicher Rückblick zeigt, daß zwischen der Programmierung und den Rechenmaschinen ein enger Zusammenhang besteht. In Wahrheit sind programmierbare Informationsverarbeitungssysteme nichts anderes als Rechner, und umgekehrt sind Rechner nichts anderes als durch ihre Programmierbarkeit vielseitig einsetzbare Informationsverarbeitungssysteme. Der programmierbare elektronische Rechner, der *Computer*, ist in diesem Sinne heute das technische System zur Informationsverarbeitung. Wir flechten an dieser Stelle einige Bemerkungen zu den Computern ein, die uns später von Nutzen sein werden. Um genau zu sein, müssen wir hinzufügen, daß hier und im folgenden «Computer» im Sinne von Digital-Rechner verstanden wird. Ein Digital-Rechner nutzt die digitale Form der Informationsverarbeitung. Daneben gibt es auch sogenannte Analog-Rechner, die mit analog dargestellter Information arbeiten. Beide Formen der Infor-

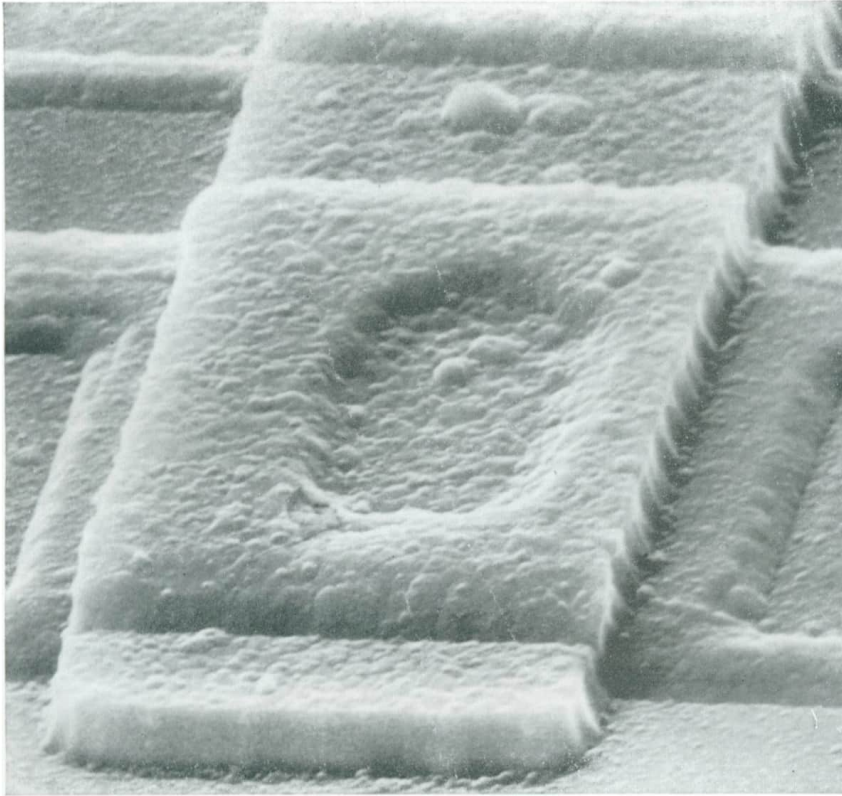
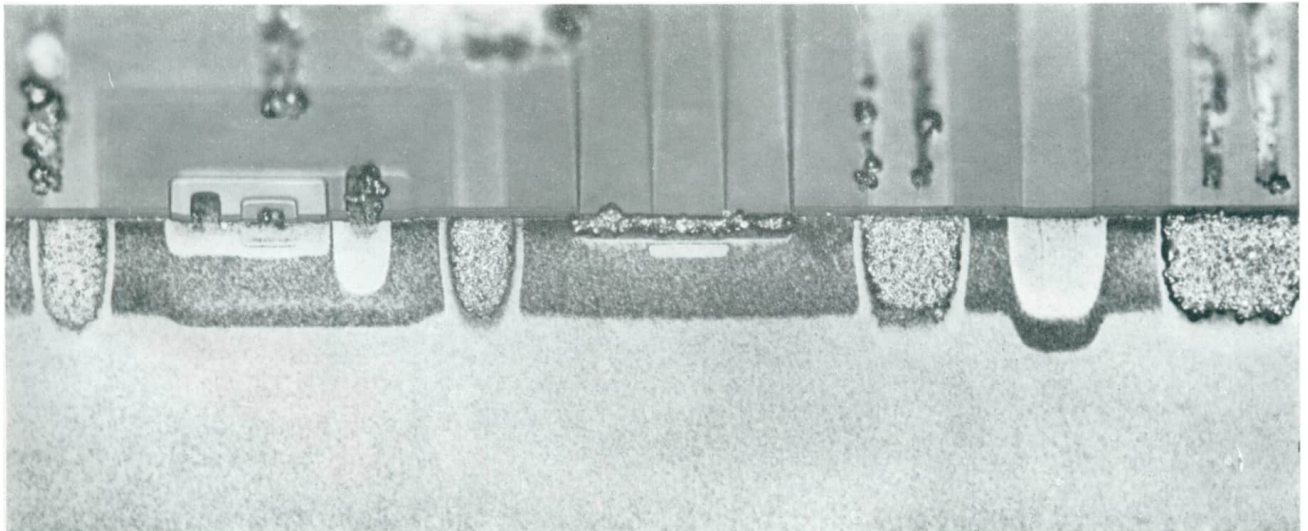


Abb. 40 Elektronenmikroskopische Aufnahme einer metallischen Leitbahn. Die an der rechten Kante sichtbaren Rillen sind beim Wegätzen des übrigen Aluminiums entstanden. Die Grube in der Mitte markiert ein Oxidfenster. An dieser Stelle hat die Leitbahn Kontakt zur darunterliegenden Siliziumschicht

Abb. 47 Aufnahme eines Schrägschliffes durch ein Siliziumplättchen mit Schaltkreisstrukturen. Die dünne, an der Kante feingeriffelte Deckschicht ist Siliziumdioxid. Die hellen und dunklen Bereiche darunter sind unterschiedlich dotiertes Silizium. Auf der Oberfläche ragen die metallischen Kontakte aus der Siliziumdioxidschicht hervor



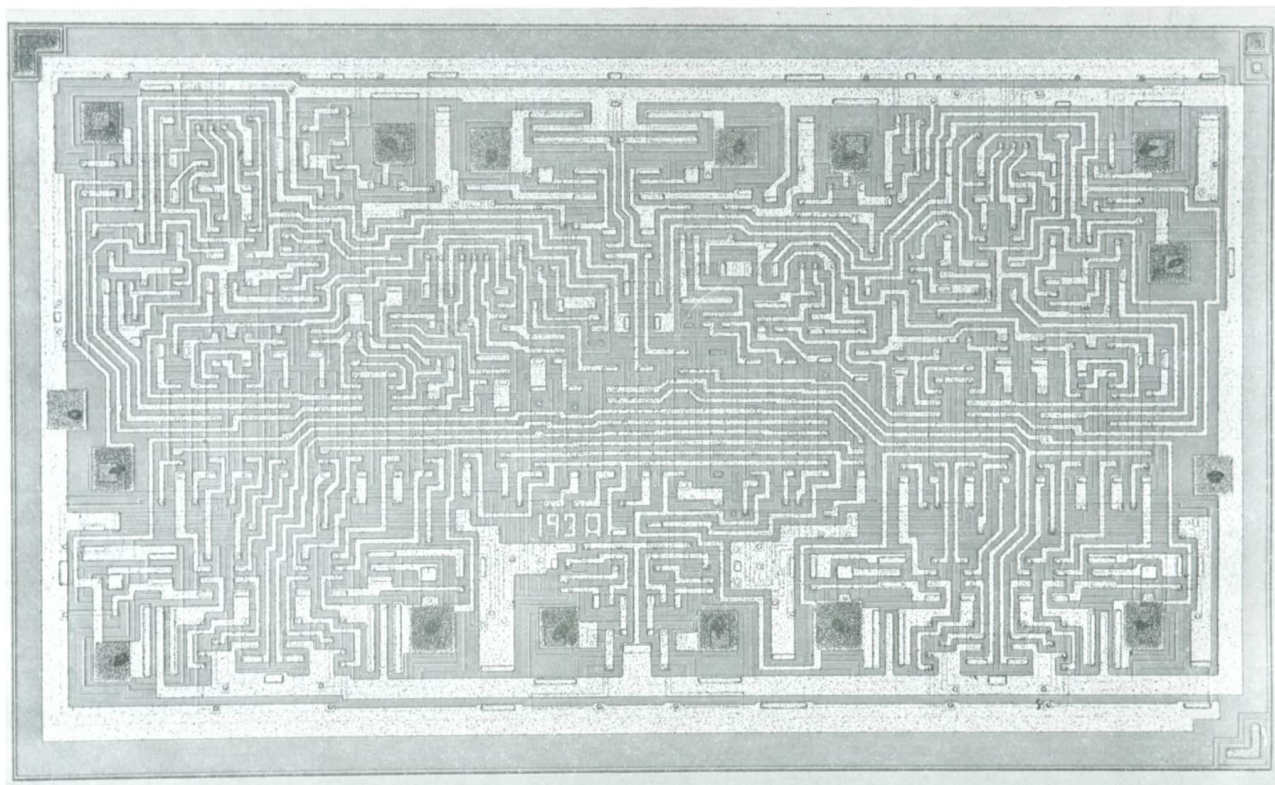
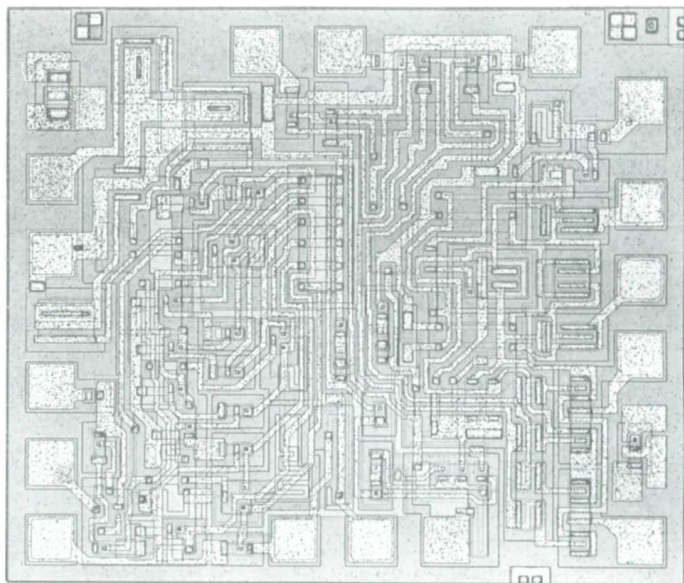
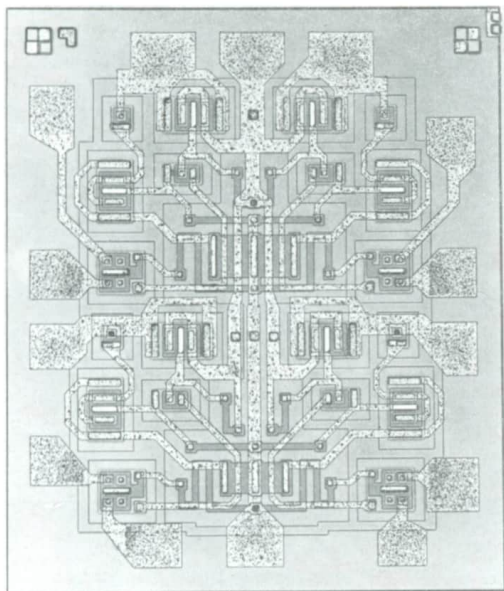
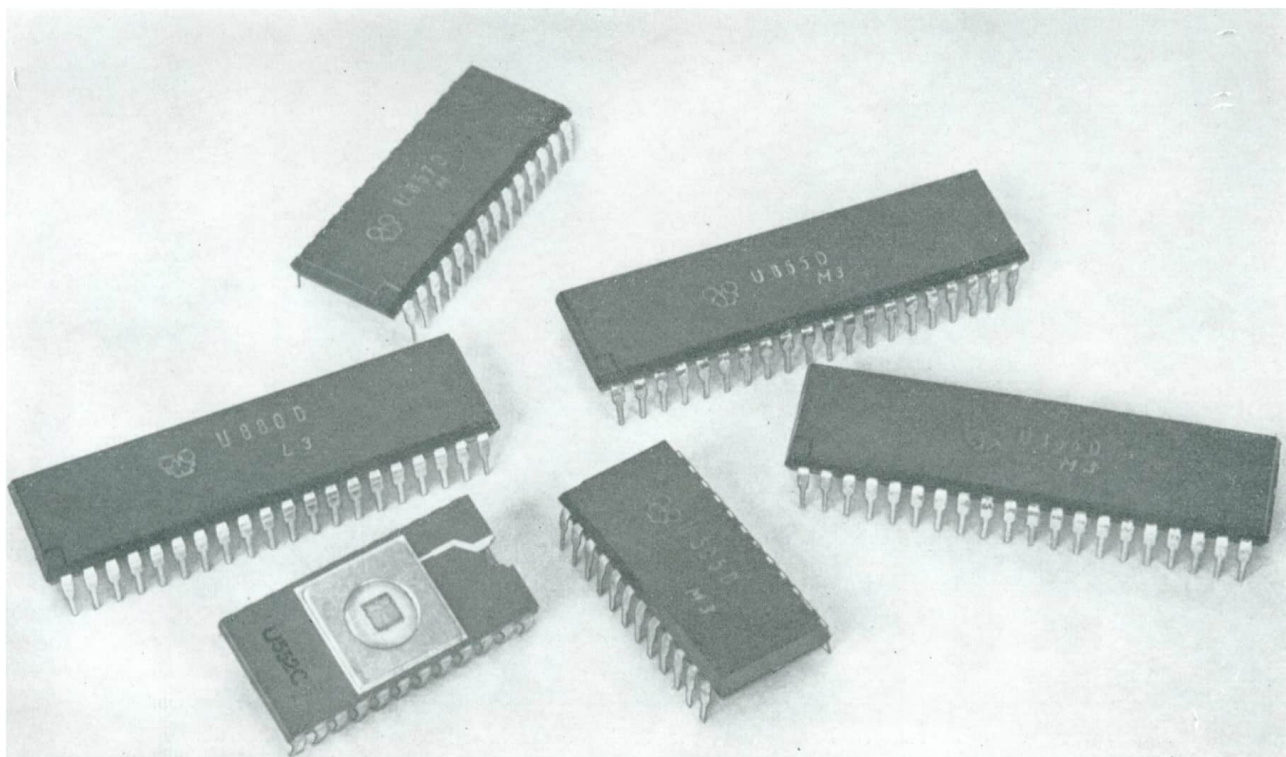


Abb. 44 Chipaufnahmen

- a) des Standard-Schaltkreises mit vier NAND-Gattern D 100 (SN7400),
- b) des Radio-Empfänger-Schaltkreises A 244,
- c) des Zählschaltkreises D193(SN74193)

Abb. 53 Kontakt zwischen einem Golddrähtchen und einer Bondinsel aus Aluminium auf einem Chip

**Abb. 54 Schaltkreise mit Plastver-
kappung U857 (Z80CTC), U880 (Z80),
U855 (Z80PIO), U856 (Z80 SIO) und
in einem Keramikgehäuse U552(1702A)**



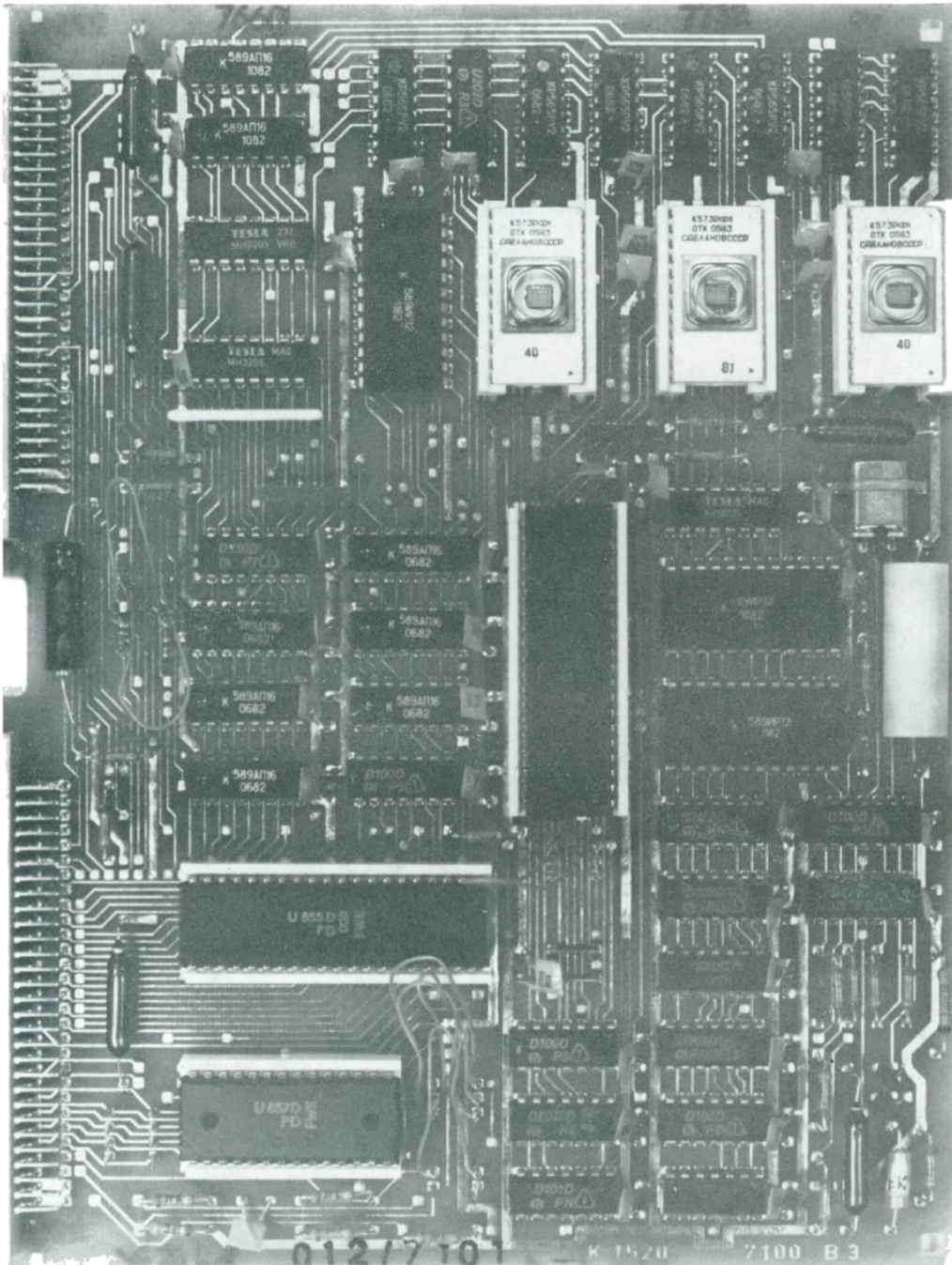


Abb. 59 Mikrorechner, wie er als zentrale Recheneinheit im Elektronenrechner K1520 des Kombinats Robotron eingesetzt wird

Tab. 9 Herkömmliche Einteilung von Computern nach Größenklassen. Heute sind die Grenzen vor allem zwischen den kleineren Computer-Typen fließend geworden und außerdem nach oben hin verschoben.

Größenklasse	Logik-Gatter in zentraler Verarbeitungseinheit (Anzahl)	Speicherkapazität (KByte)
Großrechner	über 10^5	$> 10^3$
Kleinrechner	einige 10^4	$\leq 10^3$
Minirechner	5 bis $10 \cdot 10^3$	einige 10^2
Mikrorechner	einige 10^3	einige 10

mationsverarbeitung werden in den sogenannten Hybrid-Rechnern verwendet. Die Digital-Rechner sind die weitaus verbreitetsten und universellsten. Die Leistungsfähigkeit eines solchen Computers hängt vor allem von der Größe seiner zentralen Verarbeitungseinheit, d. h. von der Anzahl der integrierten Bauelemente, die zu ihrem Aufbau verwendet wurden, und von der Größe seiner Speicher ab. In Tabelle 9 ist eine an diesen Merkmalen orientierte Einteilung der Computer aufgeführt. Durch den Einsatz des Mikroprozessors und die damit verbundene Entwicklung vor allem der kleineren Rechnertypen ist diese Einteilung aber ins Wanken geraten. Außer nach ihrer Größe unterscheidet man die Rechner nach der Art und Realisierung der verwendeten Bauelemente. Die Rechner mit Elektronenröhren bildeten die erste Generation. Mit diskreten Halbleiterbauelementen waren die Rechner der zweiten Generation ausgerüstet. Integrierte Schaltkreise auf SSI- und MSI-Niveau wurden etwa ab 1965 eingesetzt und waren charakteristisch für die Rechner der dritten Generation. Die Verwendung von LSI-Schaltkreisen etwa ab 1975, später auch von VLSI-Schalt-

kreisen, markiert den Übergang zu den Rechnern der vierten Generation. Gleichzeitig wurden die vorher ausschließlich verwendeten magnetischen Kernspeicher der Zentraleinheit zunehmend durch die schnelleren Halbleiterspeicher ersetzt. **Problemhäufigkeit und Stückzahl** Es gibt noch einen dritten Grund, weshalb die Mikroelektronik-Industrie nicht für jedes Informationsproblem den passenden Schaltkreis bereitstellen kann. Er ist von rein ökonomischer Natur. Stellen wir uns für den Moment einmal vor, die beiden vorher genannten Gründe bestünden nicht, d. h., es wäre technisch möglich, beliebig viele Bauelemente auf einem Chip zu integrieren und die Kapazität der Mikroelektronik-Industrie wäre auch ausreichend groß, um auf die Reduzierung der Typenvielfalt von Schaltkreisen durch Anwendung der Programmierung verzichten zu können. Mit den dann industriell herstellbaren, dem jeweiligen Verarbeitungsproblem angepaßten Schaltkreisen wäre allerdings in den meisten Fällen nichts gewonnen. Diese Schaltkreise wären viel zu teuer. Der Aufwand, der bei der Herstellung eines integrierten Schaltkreises getrieben werden muß, ist,

wie wir im vorangegangenen Kapitel gesehen haben, enorm hoch. Zu erschwinglichen Kosten kann ein Schaltkreis nur hergestellt werden, wenn er in einer genügend hohen Stückzahl abgesetzt werden kann. Gerade das ist aber bei Schaltkreisen, die auf einen einzigen speziellen Verwendungszweck zugeschnitten sind, in der Regel nicht der Fall. Über die Ausnahmen von der Regel, die Spezial- oder Kundenwunsch-Schaltkreise, wird weiter unten noch gesprochen werden. Unter dem Gesichtspunkt, die Schaltkreise zu möglichst niedrigen Preisen anbieten zu können, darf die Industrie also nicht darauf ausgehen, für jedes Informationsverarbeitungsproblem einen eigenen Schaltkreis im Sortiment zu haben, sondern sie muß ganz im Gegenteil versuchen, mit einem möglichst kleinen Sortiment verschiedener Schaltkreise, im Extremfall nur mit *einem* einzigen Schaltkreis, alle Kundenwünsche zu befriedigen. Dieser eine integrierte Schaltkreis zur Lösung aller Informationsprobleme, gäbe es ihn wirklich, würde in einer riesigen Stückzahl benötigt werden und könnte deshalb relativ billig hergestellt werden. «Relativ» bedeutet dabei bezogen auf seine Leistungsfähigkeit. Da diese sehr hoch sein müßte — der Schaltkreis soll immerhin beliebig komplizierte Informationsprobleme lösen —, könnte er allerdings trotz hoher Stückzahl nicht zu niedrigen Kosten angeboten werden. Bei den meisten speziellen Informationsverarbeitungsproblemen würde die sehr hohe Leistungsfähigkeit des Schaltkreises aber überhaupt nicht ausgenutzt werden. Der Kunde würde sein Geld für etwas ausgeben

müssen, das er überhaupt nur zu einem geringen Bruchteil nutzen könnte. In dem einen universellen Schaltkreistyp für alle Informationsverarbeitungsprobleme liegt also die Lösung des Problems, wie man zu billigen Schaltkreisen kommen kann, ebenso wenig wie in der ungeheuren Typenvielfalt von Spezialschaltkreisen, die jeweils *einem* Informationsverarbeitungsproblem auf den Leib geschrieben sind.

Fassen wir zusammen: Die integrierten Schaltkreise, die von der Mikroelektronik-Industrie erwartet werden, sollen Schaltungen zur Lösung der in der Praxis auftretenden Probleme der Informationsverarbeitung entweder bereits sein, oder es sollen sich aus ihnen solche Schaltungen zusammensetzen lassen. Für das Sortiment an integrierten Schaltkreisen ist einerseits bestimmend, wozu die Industrie technisch in der Lage ist, d. h. vor allem, wieviele integrierte Bauelemente sie auf einem Chip unterbringen kann. Andererseits müssen die angebotenen Schaltkreise vielseitig nutzbar, d. h. programmierbar oder zum Aufbau programmierbarer Informationsverarbeitungssysteme geeignet sein. Schließlich sollen sich die Schaltkreise in hoher Stückzahl absetzen lassen.

Kundenwunsch- und Standardschaltkreise

Das tatsächliche Angebot der Industrie an integrierten Schaltkreisen ist ein Kompromiß aus diesen Forderungen. Das Sortiment läßt sich im wesentlichen in zwei große Gruppen einteilen: Erstens die *Kundenwunsch- oder Spezialschaltkreise* und zweitens die *Standardschaltkreise*. Allerdings gibt es keine scharfe Grenze zwischen diesen beiden Gruppen,

man kennt auch *Halb-Kundenwunsch-Schaltkreise*, und manchmal ist es Ansichtssache, ob man einen bestimmten Schaltkreis zu den Spezial- oder zu den Standard-Schaltkreisen rechnet. Letzten Endes ist das auch nicht so wichtig. Wenn wir trotzdem auf diese Unterscheidung etwas ausführlicher eingehen, so geschieht das vor allem deshalb, weil sie indirekt etwas mit der mehr grundsätzlichen Frage nach einer Struktur unter den Informationsverarbeitungsproblemen und unter den Schaltungen zu ihrer Lösung zu tun hat. Primär ist die Unterscheidung in Kunden- und Standard-Schaltkreise aber daran orientiert, welche Schaltkreise von der Industrie zum einen auf spezielle Anforderung eines Kunden und zum anderen standardmäßig hergestellt werden. Ein Spezial- oder Kundenwunschschaltkreis ist in der Regel ein Schaltkreis, der ein spezielles Informationsverarbeitungsproblem, wie es bei einem Kunden vorliegt, vollständig zu lösen imstande ist. Der Begriff «Kunde» ist dabei nicht allzu wörtlich zu nehmen. Der Kunde kann ein ganzer Industriezweig sein. Ein typisches Beispiel für einen Spezialschaltkreis ist der Schaltkreis einer elektronischen Uhr, der Zeitinformationen in Form von Sekunden-, Minuten-, Stunden- und Datumsangaben erzeugt. Ebenso wie der Uhrenschaltkreis ist auch jeder andere Spezialschaltkreis nur für die Lösung eines einzigen Informationsverarbeitungsproblems einsetzbar. Er ist also extrem einseitig. Seine Einsatzbreite ist sehr gering. Das eine Problem aber, für das er zuständig ist, löst er in der denkbar besten Weise und ohne die Zuhilfenahme weiterer Schaltkreise. Das

Informationsverarbeitungsproblem darf deshalb auch nicht zu kompliziert sein. Die zu seiner Lösung erforderliche Schaltung muß sich bei dem heute technisch beherrschten Integrationsgrad noch auf *einem* Chip unterbringen lassen. Ein Spezialschaltkreis zum Schachspielen auf einem auch nur einigermaßen anspruchsvollen Niveau ist weder heute noch in absehbarer Zukunft realisierbar. Damit sich ein Spezialschaltkreis sowohl für den Hersteller als auch für den Kunden bezahlt macht, muß das Problem, für das er bestimmt ist, in exakt der gleichen Form genügend häufig auftreten. Was dabei «genügend häufig» bedeutet, hängt vom Integrationsgrad des Schaltkreises ab. Ist er hoch, rentiert sich nur eine sehr hohe Stückzahl. Ist er niedriger, reicht eine kleinere aus. Bei einem MSI-Schaltkreis kann man ganz grob 10000 Stück pro Jahr als untere Rentabilitätsgrenze ansetzen.

Das Gegenstück zu den Spezialschaltkreisen sind die Standardschaltkreise. Sie lösen in der Regel kein komplexes Informationsverarbeitungsproblem, so wie es bei einem Anwender unmittelbar auftritt, *vollständig*. Standardschaltkreise sind in der Regel Schaltkreise für häufig auftretende *Teilschritte* von Informationsverarbeitungsproblemen. Man kann sie als die standardisierten Bauteile auffassen, aus denen sich komplette Schaltungen zur Lösung von Informationsverarbeitungsproblemen aufbauen lassen. Für eine komplette Schaltung benötigt man in der Regel mehrere Standardschaltkreise. Dafür lassen sich aber auch die unterschiedlichsten Schaltungen aus ihnen zusam-

menstellen. Die Anwendungsbreite der Standardschaltkreise ist also hoch. Dementsprechend hoch ist auch die Stückzahl, in der sie abgesetzt werden können. Zu den Standardschaltkreisen gehören Schaltkreise für die elementaren logischen Operationen wie AND und OR, aber auch Schaltkreise für häufig wiederkehrende Verknüpfungen zwischen ihnen, wie zum Beispiel die Addition, das Speichern oder das Zählen. Auch Mikroprozessoren sind Standardschaltkreise. Im Unterschied zu den vorher genannten Schaltkreisen sind Mikroprozessoren aber programmierbar, d. h., die Verknüpfungen zwischen den elementaren logischen Operationen sind bei ihnen nicht ein für alle Mal fixiert, sondern können durch ein Programm, das in einem mit dem Mikroprozessor verbundenen Speicherschaltkreis untergebracht ist, in bestimmter Weise eingestellt werden. Wünscht man andere Verknüpfungen, braucht man nur das Programm im Halbleiterspeicher zu ändern. Das Prinzip der Programmierbarkeit, das beim elektronischen Rechner auf der Ebene eines aus vielen integrierten Schaltkreisen aufgebauten Systems realisiert wird, ist beim Mikroprozessor gewissermaßen auf der Ebene eines einzelnen integrierten Schaltkreises verwirklicht. Ein Mikroprozessor in Verbindung mit einem Speicherschaltkreis erfüllt wesentliche Funktionen eines programmierbaren Informationsverarbeitungssystems, d. h. eines elektronischen Rechners. Er ist die zentrale Verarbeitungseinheit eines elektronischen Rechners, gemessen an der Leistungsfähigkeit heute im allgemeinen noch eines Mikrorechners. Bei den Mikropro-

zessoren wird durch die Programmierung die zunächst offen gelassene Funktion in bestimmter Weise festgelegt. Hardware-mäßig ändert sich bei der Programmierung am Mikroprozessor nichts. Das ist aber bei einer anderen Art programmierbarer Standard-Schaltkreise der Fall — den Festwertspeichern (englisch ROMs) und den programmierbaren logischen Feldern (englisch PLAs). Auf den Chips dieser Schaltkreise sind zunächst nicht alle Transistor-Gates in der später gewünschten Weise mit den Spannungsniveaus «0» oder «1» verbunden. Die gewünschten leitenden Verbindungen werden im Nachgang zu dem bis dahin für alle Chips einheitlichen Herstellungsprozeß erzeugt. Die differenzierte Weiterverarbeitung einzelner Chipgruppen im Nachgang besorgt entweder die Industrie selbst, indem sie unter Verwendung spezieller Fotoschablonen die jeweils gewünschten Verbindungen in einem zusätzlichen Metallisierungsschritt auf die Chips aufbringt. Man spricht in diesem Zusammenhang von Maskenprogrammierung. Oder der Kunde hat die gewünschten Verbindungen selbst herzustellen. Dafür werden verschiedene Verfahren angewandt, z. B. das Wegschmelzen unerwünschter Leitbahnen durch starke Stromstöße. Festwertspeicher, die auf diese Weise vom Kunden programmiert werden können, bezeichnet man als «Programmierbare Festwertspeicher» (englisch PROMs).

Im Unterschied zu den Mikroprozessoren sind bei den ROMs, PROMs und PLAs die gewünschten leitenden Verbindungen der Transistoren nach der Programmierung im Schaltkreis Hardware-mäßig fixiert. Das Prinzip der Hardware-mäßigen Programmie-

rung wird in gewissem Sinne auch bei den sogenannten Zellen — oder Gatearrays angewandt, nur daß es bei letzteren bereits auf einer früheren Stufe des Fertigungsprozesses einsetzt und daß im Ergebnis kundenspezifische Schaltkreise in noch größerer Vielfalt hergestellt werden können, als das bei den ROMs und PLAs der Fall ist. Gatearrays — auch die Bezeichnung «Master-Slices» ist üblich — sind Chips, auf denen Schaltungselemente, sogenannte Zellen, weniger verschiedener Grundtypen regelmäßig angeordnet sind. Solche Zellen können einzelne Transistoren oder Gruppen aus Transistoren sein. Aus den Gatearray-Chips lassen sich durch Aufbringen metallischer Verbindungen Spezialschaltkreise für unterschiedliche Kundenwünsche herstellen. Meist werden zwei Metallisierungsschritte angewandt — im ersten Schritt werden die Verbindungen innerhalb der Zellen realisiert, im zweiten die zwischen den Zellen. Das Kundenspezifische bei dieser Art der Schaltkreisherstellung ist auf die beiden Metallisierungsschritte reduziert. Nur dafür muß ein spezifisches Layout entworfen werden. Der Aufwand für den Schaltkreisentwurf ist dadurch relativ gering. Die unter Verwendung von Gatearrays hergestellten Schaltkreise sind zwar einerseits Kundenwunschschaltkreise, andererseits aber auch Standardschaltkreise, nämlich insofern, als sie weitgehend standardmäßig hergestellt werden können. Durch die Gatearrays wird die auch vorher schon unscharfe Grenze zwischen den Standard- und Kundenwunschschaltkreisen noch weiter verwischt. Die Gatearrays liegen in der Mitte zwischen den beiden Schaltkreisarten. Sie vereinigen

bis zu einem gewissen Grade die Vorteile beider in sich. Man bezeichnet sie deshalb mitunter auch als *Halbkundenwunschsaltkreise*. In den letzten Jahren haben die Gatearrays zunehmend an Bedeutung gewonnen. Mit ihrer Hilfe lassen sich Spezialschaltkreise auch bei kleinen Stückzahlen (kleiner als 10Tausend) kostengünstig herstellen. Programmierbare und nichtprogrammierbare Standardschaltkreise unterscheiden sich meist auch durch ihren Integrationsgrad. Die programmierbaren Standardschaltkreise haben in der Regel ein höheres (LSI- und VLSI-)Niveau als die nichtprogrammierbaren, die meist nur MSI-Niveau erreichen oder darunterliegen. Der Grund dafür ist leicht zu verstehen. Je höher der Integrationsgrad eines Schaltkreises ist, um so spezieller ist der Schaltkreis auch. Soll er vielseitig einsetzbar sein, darf der Integrationsgrad nicht zu hoch

werden — oder der Schaltkreis muß sich programmieren lassen. Die Programmierbarkeit ist der Ausweg, der gefunden werden mußte, um Standardschaltkreise mit hohem Integrationsgrad zu ermöglichen. Bevor wir auf die soeben charakterisierten Schaltkreistypen genauer eingehen, wollen wir uns zunächst noch einmal mit der Frage ihrer Einsatzmöglichkeiten beschäftigen. Welche Schaltkreise wird man anwenden, wenn es um den Aufbau einer Schaltung zur Lösung eines bestimmten Informationsverarbeitungsproblems geht? Im Prinzip kommen dafür zunächst einmal alle fünf der oben genannten Schaltkreistypen in Frage — die Kundenwunsch- oder Spezialschaltkreise, die nichtprogrammierbaren Standardschaltkreise, die Software-mäßig programmierbaren Standardschaltkreise, insbesondere die Mikroprozessoren, die Hardware-mäßig programmierbaren Standard-

saltkreise, darunter die ROMs, PROMs und PLAs und schließlich die Gatearrays. Jede dieser fünf Lösungen hat ihre Vor- und Nachteile. In Tabelle 10 sind sie für die drei heute wohl noch am häufigsten verwendeten Lösungen zusammengestellt — den Kundenwunschsaltkreis, die Schaltung aus nichtprogrammierbaren Standardschaltkreisen und den Mikrorechner. Man erkennt, daß es von der Art des Problems und von den Anforderungen an das Verarbeitungssystem abhängt, welcher Variante der Vorzug zu geben ist. Bei Informationsverarbeitungsproblemen, die in exakt der gleichen Form genügend häufig auftreten, ist der Kundenwunschsaltkreis nicht nur die billigste, sondern in der Regel auch die platzgünstigste, energiesparendste und, sofern die gleiche Realisierungstechnik für die zur Diskussion stehenden Schaltkreise verwendet wird, auch die

Tab. 10 *Drei verschiedene Varianten zur Lösung von Informationsverarbeitungsproblemen*

Spezialsaltkreis	Schaltung aus nicht programmierbaren Standardschaltkreisen	Mikrorechner
hohe Stückzahl	niedrige Stückzahl	beliebige Stückzahl
nicht zu hohe Komplexität des Problems	geringe Komplexität des Problems	hohe Komplexität des Problems
Problem ändert sich nicht	Problem ändert sich nicht	Änderung des Problems zu erwarten
Verarbeitungsgeschwindigkeit u. U. hoch	hohe Verarbeitungsgeschwindigkeit gefordert	keine extrem hohe Verarbeitungsgeschwindigkeit notwendig
minimaler Platz und Leistungsverbrauch	Platz und Leistungsverbrauch unwesentlich	keine speziellen Anforderungen an Platz und Leistungsverbrauch
keine speziellen Anforderungen an Entwicklungszeit	mittlere Entwicklungszeit ausreichend	kurze Entwicklungszeit gefordert

schnellste Variante. Da wegen des sonst zu hohen Aufwandes bei Kundenwunschsaltkreisen aber oft langsamere Realisierungstechniken als bei Standardsaltkreisen zum Einsatz kommen, ist der Kundenwunschsaltkreis oft doch nicht die schnellste, sondern mitunter sogar die langsamste Variante. Das Informationsverarbeitungsproblem muß bei dieser Lösung einfach genug sein, um bei den heute beherrschten Integrationsgraden überhaupt mit einem einzigen integrierten Schaltkreis bewältigt werden zu können. Die Mikrorechner-Lösung, d. h. der Aufbau des Verarbeitungssystems aus Mikroprozessoren und Speichersaltkreisen, kann ihre Vorteile da entfalten, wo ein relativ umfangreiches Informationsverarbeitungsproblem gelöst werden muß, das zudem noch von Zeit zu Zeit abzuändern ist und wo das Verarbeitungssystem schnell verfügbar sein soll. Der Aufbau und die Programmierung des Mikrorechners erfordern nämlich wesentlich weniger Zeit als die Entwicklung und Herstellung eines speziellen Schaltkreises. Der Aufbau des Verarbeitungssystems aus nichtprogrammierbaren Standardsaltkreisen kommt bei einem komplexen Informationsverarbeitungsproblem in der Regel nicht in Frage. Mit Ausnahme von großen Rechnern, wo vor der Erfindung des Mikroprozessors nichts anderes übrig blieb, als nichtprogrammierbare Standardsaltkreise zu verwenden, ist diese Variante auf einfache Probleme beschränkt geblieben. Ihr Vorteil ist vor allem die hohe Verarbeitungsgeschwindigkeit. Sie wird möglich, weil bei dieser Lösung die schnellen bipolaren TTL-Saltkreise eingesetzt werden können.

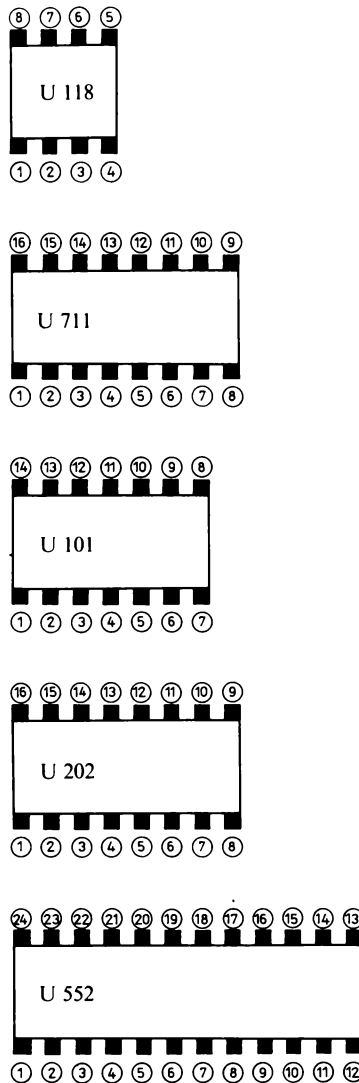


Abb. 57 Schematische Darstellung einiger Schaltkreise

Im folgenden wollen wir an Hand von Beispielen ausführlicher auf die einzelnen Schaltkreistypen eingehen. Dabei kann es sich wirklich nur um Beispiele handeln und um keine systematische, den Anspruch auf Vollständigkeit erhebende Darstellung der heute auf dem inter-

nationalen Markt erhältlichen integrierten Schaltkreise. Die Beispiele sind vorwiegend aus dem Angebot des Kombinati «Mikroelektronik» der DDR gewählt. Da, wo es möglich ist, geben wir neben der in der DDR üblichen Bezeichnung in Klammern auch die internationale Bezeichnung für die Schaltkreise an.

Wir beginnen mit den Kunden- oder Spezialsaltkreisen.

Uhrensaltkreis

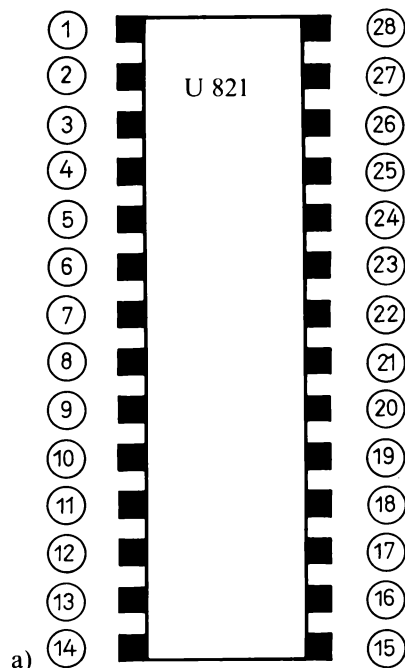
In Kapitel 4 hatten wir eine Schaltung zum Zählen bis zwei kennengelernt. Durch Zusammenfügen vieler solcher Schaltungen entsteht ein Spezialsaltkreis, wie er für eine elektronische Uhr benötigt wird. Der Schaltkreis U118 des Funkwerkes Erfurt ist ein Beispiel dafür (siehe Abb. 57). Er ist für die Verwendung in einer einfachen Armbanduhr mit analoger Anzeige bestimmt. Er besitzt acht Anschlüsse oder Pins, wie die gebräuchliche englische Bezeichnung lautet. An die beiden Pins 1 und 5 wird die Versorgungsspannung angelegt. Über die beiden Eingänge 7 und 8 wird eine mit hoher Frequenz schwingende elektrische Spannung zugeführt. Beim U118 muß die Schwingungsfrequenz gleich $2^{15} = 32768$ Hertz sein, damit an den Ausgängen 2 und 3 nach jeweils 2 Sekunden genau ein Impuls entsteht. Der Impuls bei 3 ist um eine Sekunde gegenüber dem bei 1 zeitlich versetzt, so daß pro Sekunde genau ein Impuls entsteht. Die Impulse dienen zum Antrieb eines Schrittmotors, der sich bei jedem Impuls um 6 Bogengrad, d. h. um eine Zeitsekunde, weiterdreht. Am Ausgang 6 kann eine Impulsfolge von 4096 Hertz abgenommen

werden. Wird dieser Anschluß mit einem kleinen Lautsprecher verbunden, so entsteht ein Signaltön. Pin 4 dient zum Einstellen der Uhr; jeder Impuls an diesem Eingang setzt den Schaltkreis um eine Sekunde zurück. Die zum Betrieb des Uhrenschaltkreises erforderlichen elektrischen Spannungsschwingungen werden durch einen kleinen Quarzkristall erzeugt. Quarz ist ein piezoelektrisches Material, d. h., mechanische Spannungen sind bei ihm immer mit elektrischen Spannungen verbunden. Für bestimmte Schwingungsfrequenzen, die sogenannten *Eigenfrequenzen*, hat der Quarzkristall eine besondere Vorliebe. Die Werte dieser Eigenfrequenzen hängen praktisch nur von der Größe des Kristalls und der Art seiner Schnittflächen ab. Vor allem sind sie fast völlig unabhängig von der Temperatur. Wird an zwei gegenüberliegenden Flächen des Kristalls die Spannung eines elektrischen Schwingkreises angelegt, dessen Frequenz ungefähr mit der Kristall-Eigenfrequenz übereinstimmt, so gerät der Kristall in lebhaft mechanische Schwingung. Die Frequenz dieser Schwingung ist jetzt aber nicht, wie beim elektrischen Kreis, *ungefähr* gleich der Kristall-Eigenfrequenz, sondern *exakt* gleich. Auf diese Weise werden Schwankungen der Frequenz des elektrischen Schwingkreises, wie sie unvermeidlich sind, durch den Quarzkristall ausgeglichen. Die Schwingungsfrequenz wird *stabilisiert*. In dieser Schwingungsstabilisierung besteht das Geheimnis der hohen Genauigkeit einer elektronischen Uhr. Bei einer mechanischen Uhr werden die Schwingungen des Pendels oder die Drehschwingungen der Unruhe ausgenutzt. Die ent-

sprechenden Schwingungsfrequenzen sind relativ instabil. Deshalb sind mechanische Uhren wesentlich ungenauer als die quartzesteuerten elektrischen. Uhrenschaltkreise werden heute in einer Reihe unterschiedlicher Varianten mit verschiedenen Zusatzschaltungen (Weckautomatik, Stoppuhr, Radio, Taschenrechner) hergestellt. Werden Anschlüsse für Signaltöne verschiedener Höhe vorgesehen, so kann vom Uhrenschaltkreis eine Melodie erzeugt werden.

Taschenrechnerschaltkreis

Neben den quartzesteuerten elektronischen Uhren sind es vor allem die Taschenrechner, die die Mikroelektronik populär gemacht haben. Einfache Taschenrechner für die vier Grundrechenarten und einige wenige zusätzliche Funktionen (ein Speicher, Prozentbildung) können seit längerem mit Hilfe eines einzigen Schaltkreises realisiert werden. Heute ist das auch bei komfortableren programmierbaren Taschenrechnern möglich. Zur Erläuterung der prinzipiellen Wirkungsweise eines einfachen Taschenrechnerschaltkreises wählen wir den Schaltkreis U821 des Funkwerkes Erfurt. Das Gehäuse mit den Anschlußpolen ist in Abb. 58a dargestellt. Über Pin 1 werden die Taktimpulse zugeführt. Die Taktfrequenz liegt zwischen 100 und 270 kHz. Auf die Bedeutung des Taktimpulses gehen wir im Zusammenhang mit dem Mikroprozessor später noch näher ein. Die Eingänge 14, 15 und 28 werden mit äußeren Spannungsquellen verbunden. Die Pins 3 bis 13 bezeichnet man als Digit-Ausgänge. Pin 26 ist der Eingang für Binärzahlen. Er muß mit den Pins 3 (für die Dezimalziffer 1),

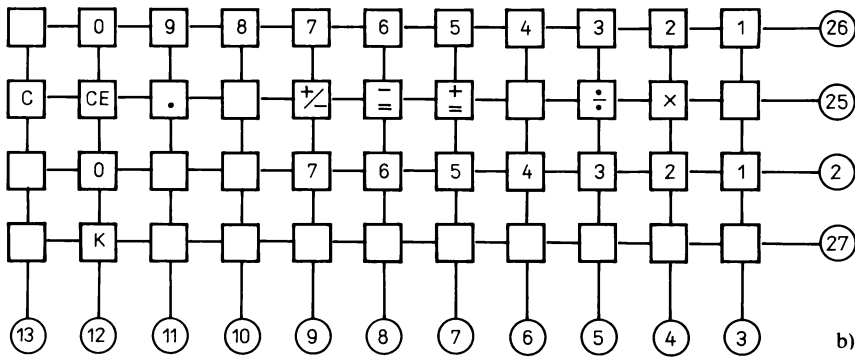


a)

Abb. 58 Zur Wirkungsweise eines einfachen Taschenrechners auf der Basis des Schaltkreises U821:

a) Anschlußbelegung des U821

4 (für 2), ... , 11 (für 9), 12 (für 0) und 13 für das Vorzeichen verbunden werden. Diese Verbindungen sind durch die entsprechenden Dezimalzifferntasten des Tastenfeldes (siehe Abb. 58b) unterbrochen. Wird die Zifferntaste «2» betätigt, so wird die Verbindung zwischen Pin 4 und dem Eingang 26 hergestellt, d. h., die Zahl «2» wird eingegeben. Über den Anschluß-Pin 25 gelangen die Rechenoperationen in den Schaltkreis, über den Anschluß 2 die Festkommastellen und über den Pin 27 der Befehl zur Konstanten-Operation. Wie die Verbindung dieser Anschlüsse mit den Digit-Ausgängen zu erfolgen hat, ist aus Abb. 58b zu entnehmen. Für die «x»-Operation muß z. B. die Verbindung zwischen

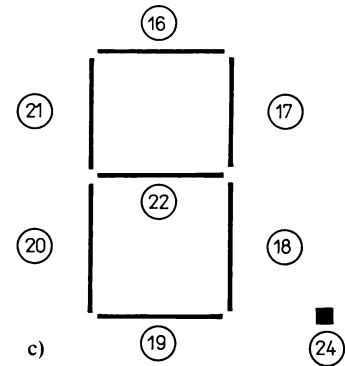


b) Tastenfeld des Rechners mit Zuordnung zu den Anschlüssen des U821

Pin 7 und Pin 25 hergestellt werden, was durch das Drücken der «x»-Taste geschieht. Das Ergebnis der Rechnung wird über die sieben Pins 16 bis 22 ausgegeben. Sie steuern die sieben Segmente der Ziffernanzeige des Rechners (siehe Abb. 58c). Über den Pin 24 wird der Dezimalpunkt der Anzeige erzeugt. Der Pin 23 ist nicht belegt.

Kameraschaltkreis

Eine Reihe von Informationsverarbeitungsproblemen, für deren Lösung Spezialschaltkreise verwendet werden können, stellen sich im Zusammenhang mit der Bedienung von Haushaltsgeräten, fotografischen Kameras oder von Geräten der Unterhaltungselektronik. Damit diese Geräte in der gewünschten Weise funktionieren, muß der Nutzer bestimmte Bedienungsoperationen ausführen. Bei einer fotografischen Kamera z. B. müssen die Helligkeit und Entfernung des Objektes gemessen und unter Beachtung bestimmter Regeln müssen Blende, Belichtungszeit und Entfernung eingestellt werden. Die mechanischen Manipulationen beim Einstellen, die normalerweise von Hand ausgeführt werden, können von einem geeignet konstruierten Mechanismus übernommen werden, der von einem kleinen Elektromotor angetrieben wird. Ein solcher Mechanismus ist in der Lage, alle möglichen Werte einzustellen. Aber er ist zunächst dumm, er weiß nicht, welche unter den gerade vorliegenden Bedingungen die richtigen sind. Dazu muß er gesteuert werden. Diese Steuerung wird von einem elektronischen Kameraschaltkreis in Verbindung mit einem Fotoempfänger ausgeführt. Durch die Fotoempfänger werden die primären optischen Informationen über Helligkeit und Abstand in analoge elektrische Signale verwandelt. Bei den komfortabelsten Kameras berechnet der Schaltkreis aus diesen Signalen und aus bestimmten Vorgaben, die der Nutzer machen muß — etwa über die Filmempfindlichkeit oder die Betriebsart — die Werte für Entfernung, Belichtung und Blende. Außerdem erzeugt er die Schaltimpulse, die über die Schrittmotoren bzw. Elektromagnete die Einstellung der richtigen Werte bewirken. Die Bedienungsoperationen an der Kamera erfolgen auf diese Weise weitgehend automatisch: Man hält die Kamera auf das Objekt, betätigt den



c) Segmentanordnung der optischen Anzeige mit Zuordnung zu den Anschlüssen

Auslöser und ist sicher, daß man ein so scharfes und gut belichtetes Bild erhält, wie man es bei der Einstellung von Belichtungszeit, Blende und Entfernung per Hand kaum hätte erreichen können. Die Automatisierung der Bedienungsoperationen wird nicht allein durch den mikroelektronischen Schaltkreis ermöglicht. Wir haben dazu auch noch einen *Sensor* für die Aufnahme der primären Information (den Fotoempfänger) und einen *Aktor* zur mechanischen Realisierung der Bedienungsoperationen benötigt (Mechanismus mit Schrittmotor oder Elektromagnet). Der elektronische Schaltkreis ist das intelligente Zentrum dieser Steuerungsanlage. Als Beispiel gehen wir kurz auf den in der Spiegelreflexkamera Practica B 200 eingesetzten Schaltkreis ein. Er enthält analoge und digitale Teilschaltkreise. Die Helligkeit des Objektes wird durch einen Gallium-Arsenid-Phosphid-Fotoempfänger gemessen. Außer der Filmempfindlichkeit muß auch die Blende vorgegeben werden. Der Schaltkreis berechnet daraus die Belichtungszeit und veranlaßt ihre Anzeige auf einer Skala durch das Aufleuchten einer

von 16 LEDs. Die Entfernungseinstellung erfolgt per Hand. Der Schaltkreis ermöglicht vier Betriebsarten der Kamera. In der Betriebsart «Automatik» wird durch Betätigung des Auslösers die Blende für die Dauer der angezeigten Belichtungszeit mit Hilfe eines Elektromagneten geöffnet.

Mit einer automatischen Steuerung, wie wir sie am Beispiel einer fotografischen Kamera besprochen haben, sind heute viele Geräte und Anlagen in Haus und Industrie ausgerüstet. Nicht immer werden dabei Spezialschaltkreise verwendet, oft kommen auch Mikrorechner zum Einsatz. In vielen Fällen wird die zu verarbeitende Eingangsinformation für den Schaltkreis auch nicht, wie beim Kameraschaltkreis, durch Sensoren aus der Umwelt aufgenommen, sondern stammt aus einem Programmspeicher, der entweder mit zum Schaltkreis gehört oder aus einem eigenen Speicherschaltkreis besteht. Im Speicher können verschiedene Programme enthalten sein, zwischen denen der Nutzer des Gerätes wählen kann. Die Bedienung des Gerätes reduziert sich in diesem Falle auf das Drücken der Taste für das gewünschte Programm; die weitere Steuerung besorgt der Schaltkreis. Diese Art der Steuerung finden wir z. B. bei Haushaltswaschmaschinen, die mit Spezialschaltkreisen oder Mikrorechnern ausgerüstet sind.

Analoge Schaltkreise

Zu den Spezialschaltkreisen gehören auch viele integrierte Analogschaltkreise, sofern sie die Eigenschaft besitzen, durch die wir einen Spezialschaltkreis charakterisiert hatten: Er muß zur vollständigen Lösung eines speziellen Informationsverar-

beitungsproblems geeignet sein. Ob die Lösung auf digitalem oder analogem Wege erfolgt, ist dabei unwesentlich. Da die Ausgangsinformationen bei einem Informationsverarbeitungsproblem, wie es unmittelbar in der Praxis auftritt, häufig zuerst in analoger Form vorliegen, sind Analogschaltkreise oft sogar besonders gut angepaßte Spezialschaltkreise. Wir erinnern außerdem an die Ausführungen über die höhere Geschwindigkeit der analogen gegenüber der digitalen Informationsverarbeitung in Kapitel 4. Typische Vertreter analoger Spezialschaltkreise sind Verstärker für elektronische Heimgeräte. Der Schaltkreis A211 des Halbleiterwerkes Frankfurt/Oder z. B. ist ein 1 Watt-Niederfrequenzverstärker zur Verstärkung von Sprach- und Musiksignalen in Rundfunk-, Fernseh-, Schallplatten oder Tonbandgeräten, der Schaltkreis A283 ein kompletter UKW-Empfänger (bis auf Filter und Spulen). Für die Decodierung der Farb- und Helligkeitssignale in einem Fernsehapparat stehen ebenfalls komplette integrierte Analogschaltkreise zur Verfügung. Ein universell einsetzbarer Verstärker, ein sogenannter Operationsverstärker, ist der Schaltkreis B 081 des Halbleiterwerkes Frankfurt/Oder (TL 081 in internationaler Bezeichnung).

Man wird dabei aber besser von einem Standard- als von einem Spezialschaltkreis sprechen. Gut geeignet sind Analogschaltkreise auch für die Stabilisierung oder Begrenzung elektrischer Spannungen und Ströme. Als wichtigen analogen Teilschaltkreis für einen AD-Wandler hatten wir in Kapitel 4 den Komparator kennengelernt. Vom Halbleiterwerk

Frankfurt/Oder wird z. B. der Komparatorschaltkreis A110 hergestellt. Auch ihn wird man besser als analogen Standardschaltkreis bezeichnen. Das Bindeglied zwischen den Analog- und Digitalschaltkreisen sind die Analog-Digital(AD)-Wandler bzw. Digital-Analog(DA)-Wandler. Sie werden heute bereits in einer großen Typenvielfalt als Standardschaltkreise angeboten. Der C520 des Halbleiterwerkes Frankfurt/Oder (AD2020) z. B. wandelt ein analoges Signal in drei binär codierte Dezimalzahlen um.

Mit den zuletzt genannten Schaltkreisen haben wir schon Beispiele für Standardschaltkreise kennengelernt. Es handelte sich dabei durchweg um *analoge* Standardschaltkreise. Wir wenden uns jetzt den *digitalen* Standardschaltkreisen zu. Zunächst geben wir Beispiele für nicht programmierbare Schaltkreise kleineren und mittleren Integrationsgrades und anschließend für hochintegrierte, programmierbare Schaltkreise.

Logik-Gatter

Typische Vertreter für Schaltkreise mit kleinem Integrationsgrad sind die verschiedenen Arten von Logik-Gattern. Meist sind gleich mehrere davon in einem Gehäuse untergebracht. Auch AND-, NAND-, OR- und NOR-Gatter mit mehr als zwei Eingängen sind üblich. Damit im Falle von AND am Ausgang eine 1 entsteht, müssen alle Eingänge mit 1 belegt sein, sonst entsteht eine 0. Im Falle von OR entsteht am Ausgang 0, wenn alle Eingänge mit 0 belegt sind, sonst 1. Die am häufigsten verwendete Logikfamilie ist die TTL-Logik. Von ihr wird z. B. in den bipolaren Schaltkreisen der D10-Serie des Halbleiterwerkes

Frankfurt/Oder (SN74) Gebrauch gemacht. Der Schaltkreis D100 (SN7400) (Abb. 44) enthält vier voneinander unabhängige NAND-Gatter mit je zwei Eingängen. Im Schaltkreis D204 (SN74H04) sind sechs voneinander unabhängige Negationsgatter zusammengefaßt.

MSI-Standardschaltkreise

Ein Schaltkreis mittlerer Integration ist der Decodierschaltkreis U711 des Funkwerkes Erfurt. Er dient der Überführung einer dreistelligen Binärzahl in die «1 aus 8» Darstellung. Wir erinnern an die Schaltung in Abb. 34, durch die eine zweistellige Binärzahl in die «1 aus 4» Darstellung umgewandelt wird. Das Gehäuse des U711 mit seinen 16 Anschlüssen ist in Abb. 57 dargestellt. Die Pins 5 bis 8 sind vier Binäreingänge, die Pins 1 bis 4 und 13 bis 16 sind die acht Ausgänge. Über die Pins 10 und 11 werden die Versorgungs- bzw. Eingangsspannungen zugeführt. Pin 12 liegt an Masse, 9 ist nicht belegt. Zu den MSI-Standardschaltkreisen gehören auch Speicher geringerer Kapazität, sogenannte *Register*. Sie werden für die kurzzeitige Speicherung kleinerer Informationsmengen verwendet. Der D181 des Halbleiterwerkes Frankfurt/Oder (SN7481), ein 16 Bit-Speicher, ist ein Beispiel dafür. Er besitzt 2 · 4 Adresseingänge, 2 Schreibeingänge und 2 Leseausgänge. Er findet als schneller Puffer- oder Zwischenspeicher Verwendung. Speicher kleineren Inhalts mit seriellem Zugriff heißen *Schieberegister*. Der Schaltkreis D191 des gleichen Herstellers (SN7491) ist ein 8 Bit-Schieberegister. Dem Speichereingang ist ein AND-Gatter mit zwei Eingängen vorgeschaltet. Die Schiebe-

register verwendet man auch als *Verzögerungsschaltkreise*. Beim Verzögerungs-Flipflop D174 (SN7474) entsteht nach dem Taktimpuls der vorher am Eingang anliegende Impuls am Ausgang. Durch die Wahl der Taktlänge kann die Verzögerungszeit eingestellt werden.

Zu den MSI-Standardschaltkreisen gehören auch die Addierschaltkreise. Vom Funkwerk Erfurt wird z. B. der Schaltkreis U101 hergestellt, durch den in zwei voneinander unabhängigen Teilschaltkreisen jeweils drei einstellige Binärzahlen addiert werden können. Dieser Schaltkreis hat 14 Anschlüsse, wie in Abb. 57 dargestellt. Die beiden Pins 1 und 4 dienen der Zuführung der Versorgungsspannung für die beiden Teilschaltkreise, Pin 11 liegt an Masse. Bei 2, 3 und 4 liegen die drei Binärzahlen an, deren Summe bei 6 (1. Stelle), 7 (2. Stelle = Übertrag) angegeben wird. Bei 14, 13 und 12 können drei weitere Binärzahlen eingegeben werden und die Summe kann bei 10 (1. Stelle) und 9 (Übertrag) abgenommen werden.

Mit diesen wenigen Beispielen für die in großer Vielfalt von der Industrie angebotenen MSI-Standardschaltkreise wollen wir es bewenden lassen und uns jetzt den hochintegrierten programmierbaren Standardschaltkreisen zuwenden. Wir konzentrieren uns dabei auf die Schaltkreise zum Aufbau eines Mikrorechners. Auf die enge Beziehung zwischen dem Mikrorechner und den hochintegrierten programmierbaren Standardschaltkreisen haben wir früher bereits hingewiesen. Wir wissen daher auch bereits, daß der Begriff «Mikrorechner» nicht zu wörtlich genommen werden darf. Ge-

meint ist ein kleines programmierbares System zur Informationsverarbeitung im allgemeinsten Sinne dieses Wortes, also nicht nur zum Ausrechnen von Zahlen, sondern auch zum Steuern von Prozeßabläufen, zum Übersetzen von Vokabeln, zum Erzeugen von Sprache oder zur Ausübung der Rolle des Gegenspielers in einem elektronischen Spielgerät.

Die Gemeinsamkeit eines Mikrorechners in diesem weiten Sinne mit einem Mikrorechner im engeren Sinne, der wirklich nur zum Ausrechnen von Zahlen bestimmt ist, besteht darin, daß sie beide einen Mikroprozessor und einen Speicher aus integrierten Schaltkreisen enthalten.

Speicherschaltkreise
halten. Wir behandeln zunächst die Speicherschaltkreise. Der Deutlichkeit halber sei darauf hingewiesen, daß hier nur von Halbleiterspeichern die Rede ist. Außer den Halbleiterspeichern gibt es andere Arten von Speichern, wie z. B. magnetische Kern-, Platten- oder Bandspeicher, Magnet-Blasen-Speicher, optische Speicher usw., die den Halbleiterspeichern an Speicherkapazität im allgemeinen weit überlegen sind.

Zum Beispiel lassen sich auf einer Plastscheibe von 30 cm Durchmesser mit einem fein gebündelten Laserstrahl 10^{10} bit Information speichern. Das entspricht mehr als 100 Büchern mit je 1000 Seiten. Die meisten dieser Speicher kommen aber bei elektronischen Rechnern nur als periphere Zusatzspeicher in Frage, und nicht als Arbeitsspeicher, die in die Funktion des Rechners unmittelbar integriert sind. Der Grund dafür ist in erster Linie die bei diesen Speichern im Vergleich zur Schnelligkeit der Verarbeitungseinheiten

des Rechners zu langsame Zugriffszeit zur gespeicherten Information. Man vergegenwärtige sich z. B. nur, daß bei einem Magnetband erst durch Vorwärts- oder Rückwärtsspulen derjenige Punkt eingestellt werden muß, von wo man den Speicherinhalt abfragen will. Auch bei besonders schnellen magnetischen Speichern, wie etwa den Ferrit-Kern- und den Blasenspeichern, liegen die Zugriffszeiten noch bei oder über einer Mikrosekunde. Damit kommen sie zwar als Arbeitsspeicher für elektronische Rechner in Frage — früher wurden sogar ausschließlich Ferrit-Kerne verwendet —, wenn man aber die Schnelligkeit der mikroelektronischen Verarbeitungseinheiten voll ausnutzen will, benötigt man Zugriffszeiten deutlich unter einer Mikrosekunde, wie sie heute von Halbleiterspeichern erreicht werden. Halbleiterspeicher sind selbst mikroelektronische Schaltungen. Die Information wird bei ihnen in der gleichen Form gespeichert, wie sie in den Verarbeitungsschaltkreisen entsteht oder benötigt wird, nämlich in Form elektrischer Impulse. Die Ein- und Ausgabe erfolgt mit der Schnelligkeit einer mikroelektronischen Schaltung. Die Zugriffszeit liegt im Bereich einiger Zehntel bis hin zu einigen Tausendstel Mikrosekunden. Halbleiterspeicher und mikroelektronische Informationsverarbeitungsschaltkreise sind einander also ideal angepaßt. Wie Halbleiterspeicher im Prinzip funktionieren, haben wir in Kapitel 4 bereits besprochen. Auch auf die besondere Stellung der *Festwertspeicher* wurde bereits hingewiesen. Danach sind Festwertspeicher gar keine richtigen Speicherschaltungen,

sondern *kombinatorische* Schaltungen, die bestimmte Eingangsimpulse zu bestimmten Ausgangsimpulsen kombinieren (siehe Kapitel 4). Als Speicherzellen können einfach FETs in matrixförmiger Anordnung verwendet werden. Die Gateelektroden der FETs sind fest mit dem «0»- oder «1»-Spannungsniveau verbunden. Ein bestimmter FET ist dementsprechend immer leitend oder nichtleitend. Daraufhin kann er abgefragt werden. Sein Zustand kann aber nicht verändert werden. Man nennt die Festwertspeicher deshalb auch ROMs — nach der englischen Bezeichnung *Read Only Memories*. Die Information, wie der Sinus eines Winkels zu berechnen ist, kann bei einem Mikrorechner in einem ROM untergebracht werden. Solange der Mikrorechner für einen bestimmten Anwendungszweck eingesetzt wird, besteht kein Anlaß, diese Information zu verändern. Wird die Versorgungsspannung an einem ROM abgeschaltet, bleibt die Information erhalten. Ohne Versorgungsspannung kann sie zwar nicht abgelesen werden, trotzdem ist sie vorhanden und sofort wieder verfügbar, wenn die Spannung eingeschaltet wird. Neben den ROMs werden Speicher für Daten benötigt, die einem mikroelektronischen Schaltkreis zur Verarbeitung zugeführt werden sollen oder die im Ergebnis der Verarbeitung entstehen, die nach ihrer Verwendung aber wieder gelöscht und durch neue ersetzt werden können. Die Speicherschaltungen, die wir in Kapitel 4 kennengelernt haben, besitzen diese Eigenschaften von vornherein. Bei ihnen kann eine durch Adressierung beliebig auswählbare Speicherzelle sowohl abgelesen als auch aufgefüllt

werden. Man nennt sie Schreib-Lese-Speicher oder RAMs — nach der englischen Bezeichnung *Random Access Memories*. Schreib-Lese-Speicher mit statischen Speicherzellen (wir erinnern an die Ausführungen dazu in Kapitel 4) heißen statische RAMs und solche mit dynamischen Zellen dynamische RAMs, kurz dRAMs. So'll ein Mikrorechner den Sinus von 33° berechnen, so müssen die « 33° » in ein RAM eingeschrieben werden. Schaltet man bei einem RAM die Versorgungsspannung ab, so geht der Speicherinhalt mehr oder weniger schnell verloren. Im gewissen Sinne eine Zwischenstellung zwischen den ROMs und RAMs nehmen die PROMs ein (*Programmable ROM*). Darunter versteht man einen von dem Anwender ein einziges Mal auffüllbaren, dann aber mit dieser Füllung für alle Zukunft weiter zu verwendenden Speicher. Bei einem EPROM (*Erasable PROM*) kann die gespeicherte Information auch wieder gelöscht werden (z. B. durch UV-Licht) und durch eine neue ersetzt werden. Bei einem E²PROM geschieht sowohl das Speichern als auch das Löschen durch starke positive oder negative elektrische Spannungsimpulse. Diese bewirken, daß an der adressierten Stelle im Speicher ein Kondensator aufgeladen oder entladen wird. Die Kondensatorplatten (das Gate und eine vergrabene Elektrode eines Feldeffekt-Transistors) sind so gut voneinander isoliert, daß die gespeicherte Ladung lange Zeit — angeben werden zehn Jahre und länger — erhalten bleibt. Ein EPROM hat also Ähnlichkeit mit einem RAM, nur daß für das Einschreiben der Information bei einem EPROM, bild-

lich gesprochen, ein härterer Stift nötig ist als beim RAM. Dafür ist die Aufzeichnung dann aber auch dauerhafter.

Zu den für den Anwender wichtigsten Merkmalen eines Speicherschaltkreises gehören die Speicherkapazität und die Zugriffszeit zu den gespeicherten Daten. Für die Anwendung in einem Mikrorechner werden Speicher von 1 KBit aufwärts eingesetzt. Sie stehen in einer relativ großen Typenvielfalt zur Verfügung — für 1, 2, 4, 8, 16, 32, 64, 128 und in jüngster Zeit auch für 256 KBit. Im Labor wurden bereits 1 MBit realisiert. Wir erinnern daran, daß 1 KBit = 2^{10} Bit = 1024 Bit, also etwa gleich 1000 Bit ist. Die Zugriffszeiten dieser Speicher liegen im Bereich zwischen 10 Nanosekunden und einer Mikrosekunde. Die japanische Firma Hitachi stellt z. B. ein 4 KBit RAM mit einer Zugriffszeit von 18 ns her. International weit verbreitet ist u. a. ein 64 KBit dynamisches RAM, das unter der Typenbezeichnung 4164 von verschiedenen Herstellern angeboten wird. 1983 wurden davon über 300 Millionen Stück verkauft, was einem Umsatz von über 3 Milliarden Mark entspricht. Das Kombinat Mikroelektronik der DDR stellt mit dem U264 ein dem 4164 vergleichbares 64 KBit dynamisches RAM her. Der U256 des Funkwerkes Erfurt ist ein 16 KBit dynamisches RAM und der U202 des gleichen Herstellers, den wir in Abb. 45 in einer Chip-Aufnahme bereits kennengelernt haben, ein 1 KBit statisches RAM. In Abb. 57 ist das Gehäuse des U202 mit den 16 Anschlüssen schematisch dargestellt. Die Pins 1, 2, 4, 5, 6, 7, 8, 14, 15, 16 dienen der Adressierung der Spei-

cherzellen. Damit können alle 2^{10} = 1024 Speicherzellen direkt adressiert werden. Über den Anschluß 3 wird dem Speicher mitgeteilt, ob geschrieben oder gelesen werden soll: 1 an diesem Eingang bedeutet lesen, 0 schreiben. Das Auslesen der Daten erfolgt über Pin 11, das Einschreiben über 12. Bei 10 wird die Versorgungsspannung angelegt, 9 liegt an Masse. Pin 13 dient zur Adressierung des Speichers in seiner Gesamtheit, sofern mehrere derartige Speicher verwendet werden (Chip-Auswahl). Die Zugriffszeit für die gespeicherten Daten beträgt 400 ns. Als Beispiele für EPROMs seien der U552 (1702A7) und der U555 (12708) des Funkwerkes Erfurt erwähnt. Die Speicherkapazität beim U552 beträgt 2 KBit, beim U555 8 KBit. Beide Speicher sind wortorganisiert, d. h., daß nicht einzelne Speicherzellen für sich, sondern nur jeweils Gruppen von Speicherzellen für ganze Datenwörter direkt adressiert werden können. Die Wortlänge beträgt 8 Bit, d. h. 1 Byte. Beim U552 (siehe Abb. 57) können also 256 Wörter von 8 Bit Länge gespeichert werden, denn $256 \text{ Bit} \cdot 8 = 2 \text{ KBit}$. Zur Adressierung sind achtstellige Adressen nötig, weil $256 = 2^8$ ist. Die Adressen werden über die Pins 1, 2, 3, 17, 18, 19, 20, 21 eingegeben. Die Pins 4, 5, 6, 7, 8, 9, 10, 11 sind zum Einschreiben und Lesen der 8-Bit-Wörter bestimmt. Über Pin 13 werden die Programmierimpulse von etwa 50 Volt eingegeben, die Löschung erfolgt durch UV-Licht. Die Anschlüsse 12, 15, 16 und 24 nehmen die verschiedenen Versorgungsspannungen auf, 22 und 23 sind Prüfeingänge und 14 dient der Chip-Auswahl. Die Zugriffszeit zu den Da-

ten beträgt beim U552 1 Mikrosekunde.

Diese Beispiele für Speicherschaltkreise zum Aufbau eines Mikrorechners mögen genügen. Wir richten jetzt unsere Aufmerksamkeit auf die zentrale Verarbeitungseinheit des **Mikroprozessoren**

Mikrorechners, den Mikroprozessor-schaltkreis oder einfach *Mikroprozessor*. Weiter vorn hatten wir schon einiges zu seiner Charakterisierung gesagt. Er ist ein Software-mäßig programmierbarer hochintegrierter Standardschaltkreis zur Ausführung logischer und arithmetischer Operationen. Das, was in einem größeren Rechner nur durch die Zusammenschaltung vieler verschiedener integrierter Schaltkreise erreicht wird, leistet der Mikroprozessor in einem Mikrorechner allein. Mikroprozessoren gibt es seit 1971, als es erstmalig der Firma «Intel» in den USA gelang, die zentrale Verarbeitungseinheit eines Mikrorechners auf einem einzigen Chip zu realisieren. Damit begann eine Entwicklung, die der Mikroelektronik erst zu ihrem Durchbruch auf der ganzen Breite der Informationsverarbeitung verholfen hat. Das Informationsverarbeitungssystem für «Überall und Jedermann» wurde möglich. Um den Mikroprozessor zu verstehen, benötigt man einige Grundkenntnisse über die Wirkungsweise eines programmierbaren elektronischen Rechners. Ein solcher Rechner arbeitet in einer zeitlichen Aufeinanderfolge von sich wiederholenden Zyklen. Jeder Zyklus ist in eine Anzahl von Takten einer bestimmten zeitlichen Länge eingeteilt. Die Takte werden durch Taktimpulse markiert, die von einem quartzesteuerten Schwingkreis, dem

sogenannten Taktgeber, erzeugt werden. Jeder Taktimpuls schaltet den Rechner in einen anderen Zustand. Nach einer gewissen Anzahl von Takten wird der Anfangszustand wieder hergestellt. Dann ist ein Zyklus durchlaufen. Die Zahl der Takte pro Zyklus ist von Rechner zu Rechner verschieden. Im folgenden wollen wir einen Zyklus in vier Takte einteilen. Unser Rechner ist also ein Viertakter. Nach jedem vierten Takt kehrt er in den Ausgangszustand zurück. Aufgaben, die man ihm stellt, löst der Rechner schrittweise. In jedem Zyklus arbeitet er sich *einen* Schritt voran. Welche Schritte er in den einzelnen Zyklen durchführen soll, weiß der Rechner, oder genauer seine zentrale Verarbeitungseinheit, die sogenannte CPU (Central Processing Unit), zunächst noch nicht. Diese Information ist in Form von Befehlen im Speicher festgehalten. Die Gesamtheit dieser Befehle bildet das Programm. Im ersten Takt jedes Zyklus muß die CPU also mit dem Speicher in Verbindung treten, um sich im zweiten Takt den Befehl holen zu können, den sie im gerade vorliegenden Zyklus ausführen soll. Im dritten und vierten Takt wird dieser Befehl ausgeführt, danach ist der erste Zyklus abgeschlossen. Dabei haben wir der Einfachheit halber angenommen, daß ein Befehl in einem einzigen Zyklus empfangen und ausgeführt werden kann. Das ist in Wirklichkeit nur bei den allereinfachsten Befehlen der Fall, für die meisten benötigt der Rechner mehrere Zyklen. Im weiteren halten wir an dieser vereinfachenden Annahme fest. Im zweiten Zyklus muß zunächst erneut die Verbindung zum Speicher hergestellt

werden, damit die CPU erfährt, was im zweiten Zyklus zu tun ist. Jedem Abholen eines Befehls aus dem Speicher in den beiden ersten Takten eines Zyklus' folgt die Ausführung des Befehls in den beiden letzten Takten. Die CPU weiß nur, was sie im gerade vorliegenden Zyklus tun soll. Was in den zukünftigen Zyklen von ihr erwartet wird, ist ihr noch unbekannt, und was sie vorher getan hat, hat sie vergessen. An die Vergangenheit erinnert sie sich nur insofern, als sie weiß, wieviele Befehle sie bereits abgearbeitet hat. Diese Erinnerung, die durch einen speziellen Teilschaltkreis, den sogenannten Befehlszähler, realisiert wird, ist für die CPU aber sehr wichtig. Aus ihr, d. h. aus der Zahl der bereits ausgeführten Befehle, werden im Befehlszähler die Adressen für diejenigen Speicherzellen erzeugt, aus denen der nächste Befehl abzuholen ist. Die Befehle sind als Binärzahlen verschlüsselt, d. h., der Speicher enthält nicht den Befehl selbst, sondern eine ihm zugeordnete Binärzahl. Diese Binärzahlen sind es, die nach Aufruf in die CPU gelangen. Dort werden sie von einem Decodierschaltkreis, dem sogenannten Befehlsdecoder, entschlüsselt. Lautet der Befehl, zwei Zahlen zueinander zu addieren, tritt die sogenannte Arithmetik-Logik-Einheit in Aktion. Sie ist ein Teilschaltkreis der CPU, in dem alle arithmetischen und logischen Operationen ausgeführt werden. Man bezeichnet sie abgekürzt auch als ALU (Arithmetik Logic Unit). Die ALU ist der eigentliche Rechenschaltkreis des Rechners. Die anderen vorher genannten Teilschaltkreise der CPU, also der Befehlszähler, die Befehlsregister und

der Befehlsdecoder, steuern zusammen mit weiteren, hier nicht genannten Teilschaltungen und dem Taktgeber die Arbeit der ALU gemäß dem im Speicher festgehaltenen Programm. Man faßt diese Schaltkreise daher unter der Bezeichnung *Steuereinheit* oder *Leitwerk* zusammen.

Ein Befehl ist die Aufforderung, an einer bestimmten Zahl eine bestimmte Operation auszuführen. Selbstverständlich ist dabei von Binärzahlen die Rede. Die Zahlen, die der Rechner verarbeiten soll, sind ebenso wie die Befehle im Speicher untergebracht. Ein Befehl kann z. B. lauten, eine Zahl aus dem Speicher in einen inneren Speicher der CPU, in ein sogenanntes Datenregister, zu transportieren. Die CPU verfügt über eine Reihe derartiger Register, um die Operanden für arithmetische oder logische Operationen und die Ergebnisse dieser Operationen kurzzeitig zu speichern. Soll zum Inhalt eines Registers eine Zahl addiert werden, so muß der Befehl erstens die Aufforderung zur Addition enthalten, zweitens aber auch die Zahl, die addiert werden soll oder wenigstens einen Hinweis, wie diese Zahl beschafft werden kann. Dieser Hinweis ist die Adresse der Zahl im Speicher. Ein Befehl besteht also aus zwei verschiedenen Teilen, der auszuführenden Operation und der Adresse des Operanden. Auch Befehle oder Teile von Befehlen werden, nachdem sie aus dem Speicher abgeholt wurden, in einem Register der CPU, dem sogenannten Befehlsregister, kurzzeitig gespeichert. Mit der ALU, der Steuereinheit und den Registern haben wir die wichtigsten Teilschaltkreise genannt, aus

denen die CPU eines programmierbaren elektronischen Rechners aufgebaut ist. Sind diese Teilschaltkreise alle in einem einzigen integrierten Schaltkreis zusammengefaßt, so ist dieser Schaltkreis genau das, was wir früher als Mikroprozessor definiert hatten. Wir wissen jetzt aber etwas genauer, wie ein Mikroprozessor aufgebaut ist und wie er arbeitet. In einem Punkt müssen wir aber unsere Beschreibung noch vervollständigen. Wir hatten bisher die Stellenzahl der Binärzahlen, mit denen der Mikroprozessor operiert und die Länge der Adressen, die von ihm erzeugt werden, offengelassen. Die Binärzahlen bezeichnet man auch als Datenwörter und die binär codierten Adressen als Adreßwörter. Ein bestimmter Mikroprozessor arbeitet immer nur mit Datenwörtern einer bestimmten Länge und mit Adreßwörtern einer bestimmten Länge. Die Länge der Daten- und Adreßwörter brauchen dabei nicht unbedingt übereinzustimmen. Mikroprozessoren unterscheiden sich u. a. durch die Länge ihrer Daten- und Adresswörter. Es gibt heute Mikroprozessoren für Datenwörter mit 4, 8, 16 und auch bereits 32 Bit Länge. Je länger die Datenwörter sind, um so größer sind auch die Zahlen, die gleichzeitig einer arithmetischen oder logischen Operation unterworfen werden können und um so schneller arbeitet demzufolge der Mikroprozessor. Die Arbeitsgeschwindigkeit hängt aber auch von der Zeit ab, die für die Ausführung eines Befehls erforderlich ist. Diese ist gleich der Dauer eines Zyklus', wenn der Befehl in einem Zyklus ausgeführt werden kann, so wie wir das bisher angenommen haben. In Mikroprozes-

soren werden aber auch Befehle verwendet, deren Ausführung mehrere Zyklen erfordert. Dann ist die Zeit zur Befehlsausführung ein Vielfaches der Zykluszeit. Da die Zykluszeit einige. in unserem Beispiel vier Taktlängen beträgt, kann man aus der Taktfrequenz auf die Schnelligkeit des Rechners schließen. Je höher die Taktfrequenz ist, um so schneller ist der Rechner im allgemeinen. Durch die Länge der Adreßwörter ist festgelegt, wie viele Speicherzellen bzw. Gruppen von Speicherzellen direkt adressiert werden können. Bei einer Länge von 8 Bit sind es $2^8 = 256$, bei 16 Bit $2^{16} = 65536$. Spitzenwerte liegen heute bei und bereits über 20 Bit. Mit 20 Bit können über eine Million Speicherzellen direkt adressiert werden. Die Befehle sind als Datenwörter verschlüsselt, auch Befehle aus mehreren Datenwörtern sind möglich. Mikroprozessoren werden heute von vielen Firmen in einer großen Typenvielfalt hergestellt. Man kann sie in verschiedene Leistungsklassen einteilen. Ein Mikroprozessor der unteren Leistungsklasse ist der U808 des Funkwerkes Erfurt. Die Länge der Datenwörter beträgt 8 Bit, die der Adreßwörter 14. Sein Befehlsvorrat umfaßt 48 Befehle. Die Zeit zur Ausführung der Befehle liegt zwischen 12 und 44 Mikrosekunden. Der U808 ist in p-Kanal-MOS-Technologie hergestellt. Ein anderer Mikroprozessor des gleichen Herstellers, der U880 (Z80) gehört zur mittleren Leistungsklasse. Er arbeitet mit 8 Bit Daten- und 16 Bit-Adreßwörtern und verfügt über 158 Befehle. Die Befehlsabarbeitungszeiten liegen bei einigen, minimal bei 1,6 Mikrosekunden. In Abb. 46 ist eine vergrößerte

berte fotografische Aufnahme des Chips des U880 zu sehen. Der Schaltkreis ist in einem 40poligen Plastikgehäuse verpackt. Der komplette Baustein ist in Abb. 54 dargestellt. Über 16 Anschlußpins werden die Adressen für die externen Speicher ausgegeben, acht Pins dienen der Datenein- und -ausgabe, ein Anschluß der Zuführung der Betriebsspannung (5 Volt), ein weiterer Anschluß der Verbindung mit einem äußeren Taktgeber (von 2,5, maximal 6 MHz). Über weitere Pins werden dem Mikroprozessor bestimmte Bedienungsanweisungen übermittelt, z. B. die Aufforderung, die Abarbeitung eines Programms zu unterbrechen (Interrupt-Anweisung). Andere Pins dienen der Ausgabe von Hinweisen über den Zustand des Mikroprozessors. Wir geben einige Beispiele für Befehle des U880:

- Befehle, durch die ein 8Bit- bzw. 16Bit-Wort aus einem externen Speicher in eines der internen Speicherregister des Mikroprozessors umgeladen wird;
 - Registeraustauschbefehle, durch die die Inhalte von zwei Speicherregistern des Mikroprozessors ausgetauscht werden;
 - Befehle zur Ausführung arithmetischer oder logischer Operationen mit zwei 8Bit- oder zwei 16Bit-Wörtern;
 - Sprungbefehle, durch die die Reihenfolge der Befehlsabarbeitung beeinflußt werden kann;
 - Ein- und Ausgabebefehle.
- Die höhere Geschwindigkeit des U880 gegenüber dem U808 ist vor allem durch den Übergang von der p-Kanal-Technik zur n-Kanal-Technik ermöglicht worden. Wir erinnern in diesem Zusammenhang an

die Ausführungen über die höhere Beweglichkeit der Elektronen des Siliziums gegenüber den Löchern in Kapitel 3.

Ein Beispiel für einen Mikroprozessor der höheren Leistungsklasse ist der 8086 der Firma Intel aus den USA. Er verarbeitet 16Bit-Datenwörter, die Länge eines Adreßwortes beträgt ebenfalls 16 Bit. Für die Addition von zwei 8Bit-Zahlen benötigt der 8086 0,6 μ s bei einer Taktfrequenz von 5 MHz. Andere leistungsfähige 16Bit-Mikroprozessoren sind der 68000 von Motorola und der Z8000 der Firma Zilog. Auch das Funkwerk Erfurt bietet mit dem U8000 neuerdings einen 16Bit-Mikroprozessor an. Als erste 32Bit-Mikroprozessoren sind der iAPX432 von Intel und der Z80000 von Zilog auf dem Markt.

Mikrorechner

Mikroprozessoren können zum Aufbau von programmierbaren Rechnern unterschiedlicher Leistungsfähigkeit verwendet werden. Für größere Rechner benötigt man mehrere Mikroprozessoren. Nimmt man nur einen, erhält man einen Mikrorechner, der im Falle eines 16Bit-Mikroprozessors bereits an einen Minirechner heranreicht. Der Aufbau eines Mikrorechners soll nun etwas detaillierter dargestellt werden. Außer dem Mikroprozessor besitzt der Mikrorechner einen Speicher für die Daten und das Programm. Der Speicher kann aus einem oder mehreren Speicherschaltkreisen bestehen. Dazu kommen spezielle Schaltkreise für die Ein- und Ausgabe von Daten und für die Übermittlung von Steuer- und Zustandsinformationen. Man nennt diese Schaltkreise PIOs und SIOs

(Parallel Input/Output bzw. Serial Input/Output). Schließlich wird ein Taktgeber benötigt, sofern er nicht bereits im Mikroprozessor enthalten ist. Ein Mikrorechner kann also aus wenigen integrierten Schaltkreisen zusammengestellt werden (siehe Abb. 59, S. XII). Dafür benötigt man zwar einen Spezialisten, aber es muß nicht unbedingt, wie bei anderen elektronischen Rechnern, die Computerindustrie bemüht werden. Den Mikrorechner — genauer einen von der eben beschriebenen Art — kann man deshalb auch als Rechner charakterisieren, der ohne den Umweg über die Computerindustrie von der Mikroelektronik-Industrie im wesentlichen selbst hergestellt werden kann, abgesehen von gewissen Montagearbeiten, die vom Anwender auszuführen sind. Für den Datenspeicher eines Mikrorechners kommen nur RAMs in Frage, weil Daten nicht nur gelesen, sondern auch neu eingeschrieben werden müssen. Als Programmspeicher können ROMs, PROMs oder EPROMs verwendet werden, weil das Programm nicht oder nur relativ selten zu verändern ist und außerdem beim Abschalten des Rechners nicht verloren gehen soll. Nicht selten kommen aber auch RAMs zum Einsatz. Im Prinzip reicht ein einziges RAM aus, um den gesamten Speicher für Daten und Befehle bei einem Mikrorechner zu realisieren. Allerdings muß dann ein magnetischer Speicher im Hintergrund stehen, der garantiert, daß beim Abschalten des Rechners das Programm erhalten bleibt. Das Zusammenspiel der Speicher und der Ein- und Ausgabeeinheit mit dem Mikroprozessor geschieht über Sam-

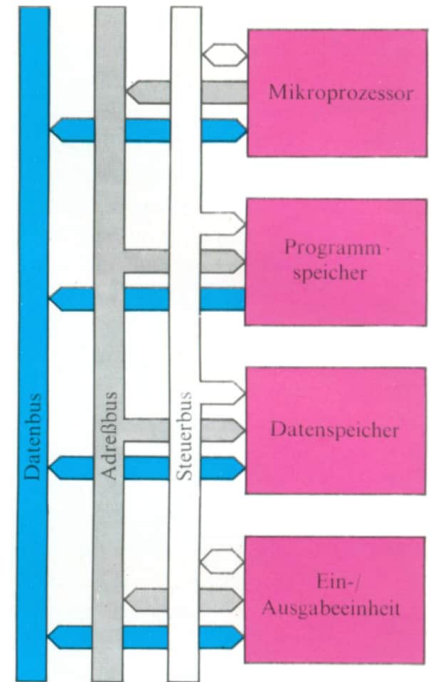


Abb. 60 Informationsfluß in einem Mikrorechner mit drei Bussen

melleitungen, sogenannte *Busse*, und besteht im Austausch von Informationen in Form von Steuersignalen, Daten und Adressen. Die Struktur dieser Busse kann unterschiedlich sein. Die meisten Mikrorechner besitzen heute für die genannten drei Arten von Information getrennte Sammelleitungen — den *Steuerbus*, den *Datenbus* und den *Adreßbus*. Der Informationsfluß in einem Mikrorechner mit drei getrennten Bussen ist in Abb. 60 schematisch dargestellt.

Am Beispiel der Addition zweier 8-stelliger Binärzahlen wollen wir jetzt verfolgen, wie ein Mikrorechner arbeitet. Wir haben dabei keinen bestimmten Mikrorechner vor Augen, und werden auch eine Reihe von Vereinfachungen an den tat-

sächlichen Abläufen in einem solchen Rechner vornehmen. Es geht uns um die Erläuterung des *Prinzips*. Immerhin wollen wir unseren Mikrorechner soweit konkretisieren, daß wir die Länge der Datenwörter zu 8 Bit, die der Adreßwörter zu 16 Bit, den Datenspeicher als RAM und den Programmspeicher als PROM annehmen. Daten- und Adreßbus seien wie in Abb. 58 voneinander getrennt. Die erste der zu addierenden Binärzahlen werde in die erste Achtergruppe des RAM eingeschrieben, die zweite Binärzahl in die zweite Achtergruppe. Das Ergebnis erwarten wir in der dritten. Ein Übertrag möge nicht auftreten. Für das Programm benötigen wir drei Befehle: Erstens den Befehl, den Inhalt der ersten Achtergruppe des RAM in ein bestimmtes Register des Mikroprozessors zu transportieren, zweitens den Befehl, die Zahl aus der zweiten Achterzellengruppe zum Inhalt dieses Registers zu addieren und das Ergebnis in ihm zu speichern, drittens den Befehl, den Inhalt dieses Registers in die dritte Achterzellengruppe des RAM umzuladen. Diese drei Befehle, d. h. jeweils eine verschlüsselte Operation und die Adresse des Operanden, speichern wir der Reihe nach im PROM ab. Pro Befehl reichen drei 8Bit-Wörter, d. h. 3 Byte, aus, eines für die Operation und zwei für die Adresse des Operanden. Das Programm kann also in den Achtergruppen 1 bis 9 untergebracht werden, wobei der erste Befehl in den Gruppen 1, 2, 3, der zweite in den Gruppen 4, 5, 6 und der dritte in den Gruppen 7, 8, 9 gespeichert werden soll. Wird der Mikrorechner eingeschaltet, stellt der erste Taktimpuls den Befehls-

zähler auf 1. Die Adresse des ersten Befehls, d. h. die der ersten drei Achterzellengruppe des PROM, wird gebildet. Gleichzeitig werden Mikroprozessor und PROM an den Adreßbus gekoppelt. Der Mikroprozessor gibt die gebildete Adresse über den Adreßbus an das PROM weiter. Durch die Adreßimpulse wird die Verbindung der Leseausgänge der adressierten Speicherzellen zum Datenbus hin geöffnet. Der zweite Taktimpuls entkoppelt den Mikroprozessor vom Adreßbus und schließt ihn an den Datenbus an. Der Inhalt der adressierten Speicherzellen des PROM, d. h. der Achtergruppen 1, 2 und 3, die den ersten Befehl enthalten, fließt über den Datenbus in den Mikroprozessor. Er wird dort im Befehlsregister gespeichert und im Befehlsdecoder entschlüsselt. Die Ausführung des Befehls kann beginnen. Sie wird durch den 3. Taktimpuls eingeleitet, der den Mikroprozessor zunächst wieder vom Datenbus entkoppelt und an den Adreßbus ankoppelt. Außerdem wird das RAM mit dem Adreßbus verbunden, während das PROM abgekoppelt wird. Der Adreßimpuls aus dem Mikroprozessor gelangt also über den Adreßbus in das RAM und öffnet die Leseausgänge der adressierten Speicherzellen, also der ersten Achtergruppe des RAM, zum Datenbus hin. Der 4. Taktimpuls schaltet den Mikroprozessor vom Adreßbus zum Datenbus um. Der Inhalt der adressierten Speicherzellen des RAM, also die erste Binärzahl, fließt danach über den Datenbus in den Mikroprozessor. Dort wird sie in demjenigen Register untergebracht, das durch den Operandenteil des Befehls bezeichnet wurde.

Damit ist der erste Befehl ausgeführt und der erste Zyklus abgeschlossen. Nachzutragen ist, daß der 4. Taktimpuls den Befehlszähler um 1 erhöht, d. h. von 1 auf 2 gestellt hat. Dadurch ist auch die Bildung der Adresse für den zweiten Befehl veranlaßt worden. Im ersten Takt des zweiten Zyklus wird diese Adresse vom Mikroprozessor über den Adreßbus an das PROM weitergegeben. Der Inhalt der adressierten Zellen, d. h. der zweite Befehl, fließt über den Datenbus in den Mikroprozessor und wird dort decodiert. Es schließt sich die Ausführung des zweiten Befehls an. Wir können uns die Beschreibung des weiteren Ablaufes im Mikrorechner schenken. Das Spiel wiederholt sich, bis das ganze Programm abgearbeitet ist. Wird an den Mikrorechner mittels eines PIO- oder SIO-Schaltkreises ein Drucker oder ein anderes Ausgabegerät angeschlossen, kann durch zusätzliche Befehle die Ausgabe des Ergebnisses veranlaßt werden. Wie bei allen anderen Befehlen erfolgt die Ausführung über den Mikroprozessor. Das ist in gewissen Fällen für die Datenausgabe zu langsam. Manche Mikrorechner verfügen deshalb auch über die Möglichkeit zur schnelleren Ausgabe von Daten aus dem Datenspeicher unter Umgehung des Mikroprozessors. Dazu ist ein spezieller Schaltkreis, ein sogenannter DMA (Direct Memory Access)-Schaltkreis erforderlich. Über ihn ist auch eine schnelle Dateneingabe möglich. Wir machen den Leser noch auf eine Eigenschaft unseres Mikrorechners aufmerksam, die heute alle elektronischen Rechner besitzen, nämlich die Gleichbehandlung von Befehlen

und Daten. Befehle werden, wie wir gesehen haben, durch Datenwörter dargestellt. Sie fließen über den gleichen Bus wie die Daten. Daraus ergeben sich wichtige Konsequenzen, z. B. kann mit Befehlen gerechnet werden und Programme können sich gegenseitig verändern. Der Erfinder dieses Prinzips ist J. von Neumann. Deshalb nennt man die heutigen elektronischen Rechner auch «von-Neumann-Rechner». Man meint damit aber meist nicht nur die Gleichbehandlung von Befehlen und Daten, sondern auch das konsequente zeitliche Nacheinander der Rechner-Operationen — zu einem bestimmten Zeitpunkt wird immer nur *eine* Operation an *einem* Operanden ausgeführt.

Nach diesen Ausführungen zum Aufbau und zur Arbeitsweise eines Mikrorechners kommen wir jetzt auf seine Anwendung zu sprechen. Sie beginnt mit der Programmierung. Der Aufwand für die Programmierung eines Mikrorechners von der oben beschriebenen Art ist relativ hoch, viel höher, als man das von einem programmierbaren Taschen- oder Tischrechner her gewohnt ist. Bei einem solchen Mikrorechner sind selbst für ein kleines Programm Zeiten von 30 Minuten und mehr zur Programmierung eines Befehls nicht zu hoch gegriffen. Im Durchschnitt kann man damit rechnen, daß einschließlich aller Vor- und Nacharbeiten von einem Mann pro Tag fünf bis zehn Befehle programmiert werden können. Der Grund dafür ist vor allem, daß diese Mikrorechner in Maschinensprache programmiert werden müssen. Das bedeutet, daß bei der Programmierung nur Befehle verwendet werden dürfen,

die im Befehlsvorrat des Mikroprozessors vorhanden sind und von ihm unmittelbar ausgeführt werden können. Wie wir bei der Behandlung des Mikroprozessors gesehen haben, beziehen sich diese Befehle grundsätzlich nur auf Binärzahlen. Während z. B. für die Addition zweier Binärzahlen ein Befehl vorgesehen ist, gibt es keinen für die Addition zweier Dezimalzahlen. Bereits zur Lösung einer solch einfachen Aufgabe muß ein kleines Programm mit mehreren Schritten geschrieben werden. Ein Programm ist auch erforderlich, wenn zwei Binärzahlen miteinander multipliziert werden sollen. Bei der Programmierung größerer Rechner können höhere Sprachen als die Maschinensprache, sogenannte Anwender-orientierte Sprachen, verwendet werden. Diese Rechner besitzen ein intern gespeichertes Übersetzungsprogramm, einen sogenannten *Compiler*, mit dem sie die höhere Programmiersprache in die auch ihnen unmittelbar verständliche Maschinensprache übersetzen. Das Übersetzen kostet Zeit, wenn nicht zusätzliche, zeitlich parallel zueinander arbeitende Schaltkreise eingesetzt werden. Die Rechner werden also entweder umfangreicher und teurer oder sie bleiben langsamer, als sie bei Programmierung in Maschinensprache sein könnten, oder es ist beides der Fall. Das ist der Preis, der für ihre einfache Programmierbarkeit gezahlt werden muß. Bei den Mikrorechnern muß man heute zwei Kategorien unterscheiden, eine, von der bisher die Rede war, bei der man bewußt darauf verzichtet, diesen Preis zu zahlen, und eine, bei der man den Preis zahlt. Zur zweiten Kategorie gehören Per-

sonal-, Büro- und Heimcomputer. Bei der ersten Kategorie nimmt man die umständliche Programmierung in Kauf und gewinnt dafür eine hohe Arbeitsgeschwindigkeit und einen niedrigen Preis. Daraus ergeben sich aber spezifische Gesichtspunkte für den Einsatz dieser Mikrorechner. Einen solchen Mikrorechner wird man nicht für die Lösung häufig wechselnder mathematischer Probleme einsetzen, sondern sinnvollerweise dort, wo sich das zu bearbeitende Problem relativ selten oder überhaupt nicht ändert. Solche Aufgaben gibt es in großer Zahl und Vielgestaltigkeit. In Tabelle 11 sind typische Einsatzfälle, vor allem für diese Kategorie von Mikrorechnern, zusammengestellt. Auf die Möglichkeit der automatischen Steuerung von Maschinen und Anlagen in der Industrie durch den Einsatz mikroelektronischer Informationstechnik haben wir schon mehrfach hingewiesen. Mikrorechner sind für diesen Zweck besonders gut geeignet. Durch den Einsatz von Mikrorechnern werden Maschinen oder Anlagen zu Automaten. Beispiele dafür sind numerisch gesteuerte Werkzeugmaschinen, automatische Zieh- und Walzstraßen, Setzautomaten in der Druckindustrie und nicht zuletzt Industrieroboter. Auf ein Beispiel unter vielen, den Setzautomaten, wollen wir näher eingehen. Die Vorstufe dafür ist die elektronische Schreibmaschine. Soll eine Textseite im Mikrorechner gespeichert werden, so benötigt man einen Datenspeicher von ca. 2 KByte. Die Eingabe der Daten, d. h. das Einschreiben des Textes in den Speicher, erfolgt über die Tastatur der Schreibmaschine. Außer dem Text müssen

Tab. 11 Mikrorechner-Anwendungen

Industrieanlagen	Numerisch gesteuerte Werkzeugmaschinen, Prüf-, Steuerungs- und Überwachungssysteme, Fließbänder, Taktstraßen, rechnende Waagen, Systeme zur Energieoptimierung in Versorgungseinrichtungen, Setzmaschinen, Kopierautomaten, Industrieroboter
Gerätebau	Registrierkassen, Fahrkartenautomaten, Aufzugssteuerungen, Tanksäulendurchflußmessung und Abrechnung, Radar, Überwachungssysteme für Umweltschutz, wissenschaftliche Geräte zur Meßwert- erfassung und -verarbeitung
Datenverarbeitungs- und Bürotechnik	Periphere Anlagen von Großrechnern, intelligente Terminals, programmierbare wissenschaftliche Tisch- rechner, elektronische Schreibmaschinen, Bürocom- puter, Personalcomputer
Nachrichtentechnik	Fernsprechwählautomaten, Meß-, Regel-, Überwa- chungssysteme
Verkehrstechnik	Zündung, Geschwindigkeitskontrolle, Kfz-Diagnose, Automatiksteuerung, Blockierschutz, Einspritz- systeme, Kfz-Radar, Bordrechner, mikrorechner- gesteuerte Ampelregelung, Straßenbahndispatcher- system
Medizintechnik	Patientenüberwachung, Röntgengerätesteuerung, EEG- und EKG-Auswertung, Analyseautomaten, Strahlendosimetrie, psychologische Testsysteme, Computer-Tomografie
Konsumgüter	Waschmaschinen, Kochherde, Nähmaschinen, Ge- schirrspülautomaten, Rundfunk- und Fernsehgeräte, Bildschirmspiele, TV-Informationssysteme, Spiel- automaten, elektronische Kameras, Heimcomputer

auch Angaben eingegeben werden, wie der Text auf dem Papier ange- ordnet sein soll, wieviel Rand ge- wünscht wird, ob ein- oder zweizeilig geschrieben werden soll usw. Der ge- speicherte Text kann danach, so oft man möchte, im gewünschten For- mat ausgegeben werden, entweder auf Papier über die Schreibmaschine selbst oder über ein Bildschirmgerät.

Beim Ausdrucken auf Papier wird das Anschlagen der Tasten durch kleine Elektromagnete bewirkt, die durch den Mikrorechner gesteuert werden. Wird nach der ersten Aus- gabe festgestellt, daß beim Einschrei- ben in den Speicher Tippfehler ge- macht wurden, kann leicht eine Kor- rektur erfolgen, indem der fehler- hafte Speicherinhalt durch den rich-

tigen überschrieben wird. Von der elektronischen Schreibmaschine zur rechnergesteuerten Anlage für die Herstellung des Satzes in der Druck- industrie ist es vom Prinzip her nur ein kleiner Schritt. Beim Druck mit Lettern übernimmt der Rechner die Steuerung der Setzmaschine, in der die Buchstabenmatrizen zur Guß- form für den Bleisatz zusammenge- fügt werden. Beim Offset-Druck steuert der Rechner die fotomecha- nische Herstellung des Lichtsatzes, d. h. der fotografischen Filme, die als Kopiervorlagen für die Druck- form (in der Regel eine beschichtete Metallplatte) dienen. Mit Hilfe des rechnergesteuerten Lichtsatzes und des Offset-Verfahrens ist es gelungen, die technische Herstellung von Bü- chern vollständig zu automatisieren, d. h. eine Anlage zu entwickeln, in die man am Anfang das Papier und die anderen notwendigen Materia- lien eingibt und aus der am Ende das fertige Buch in versandfähiger Ver- packung herauskommt. Der Text des Buches wird in einen mit dem Rechner verbundenen Speicher einge- geben. Soll ein anderes Buch ge- druckt werden, braucht man nur den Speicherinhalt zu ändern. Noch sind solche Bücher äußerlich sehr uni- form. Aber auch hier zeichnen sich Lösungen ab, etwa durch die Digita- lisierung der Schrift, die die Erzeu- gung beliebiger Schriftformen mög- lich macht.

Die Automatisierung der technischen Herstellung von Büchern vermittelt einen Eindruck davon, welche Ra- tionalisierungseffekte durch den Ein- satz von Mikrorechnern erzielt werden können. Bei den meisten der in Tabelle 11 aufgeführten Anwen- dungen könnte anstelle eines Mikro-

rechners auch ein integrierter Spezialschaltkreis oder eine aus kleinen Schaltkreisen zusammengesetzte größere Schaltung verwendet werden. Auf die Besonderheiten der verschiedenen Lösungen sind wir früher bereits eingegangen (siehe Tabelle 10). Die Vorteile des Mikrorechners gegenüber den anderen Lösungen sind seine schnelle Verfügbarkeit, seine Fähigkeit zur Lösung relativ komplizierter Aufgaben und seine Anpassungsfähigkeit an unterschiedliche Probleme, die ungeachtet der relativ umständlichen Programmierung im Gegensatz zu Spezial- und zusammengesetzten Standardschaltungen eben doch vorhanden ist. Dazu kommt, daß diese Vorteile für einen relativ niedrigen Preis zu haben sind. Die Hardware beim Mikrorechner ist jedenfalls relativ billig. Das wird auch durch die folgende Überlegung verdeutlicht. Ein Befehl im Programm eines Mikrorechners vereinigt in sich in der Regel eine größere Anzahl elementarer logischer Operationen. Wie wir wissen, sind die Befehle als Binärzahlen verschlüsselt gespeichert. Pro logischer Operation beträgt der Speicherbedarf, gemittelt über alle Befehle eines typischen Mikroprozessors, etwa 10 Bit. Hat man ein 8KBit-ROM, so lassen sich also etwa 800 logische Operationen in ihm unterbringen. Das entspricht 800 Logik-Gattern oder, wenn wir davon ausgehen, daß in einem MSI-Standardschaltkreis zehn Gatter vereinigt sind, 80 Standardschaltkreisen. Der Mikroprozessor ersetzt also zusammen mit dem 8KBit-ROM etwa 80 MSI-Standardschaltkreise. Der Aufwand an Hardware, gemessen an der Anzahl integrierter

Schaltkreise, ist also bei der Mikrorechnerlösung für ein Informationsverarbeitungssystem wesentlich kleiner als bei der Verwendung von Standardschaltkreisen. Mikroprozessoren in Verbindung mit Speicher- und anderen Schaltkreisen sind heute nicht nur eine Alternative zu Spezial- und MSI-Standardschaltkreisen, sondern zunehmend auch zu Minirechnern. Zahlreiche Aufgaben in Industrie und Verwaltung, die bisher nur dem Minirechner zugänglich waren, können mit Hilfe von 16Bit-Mikroprozessoren wesentlich effektiver gelöst werden. Deshalb hat der 16Bit-Mikroprozessor seit seinem ersten Erscheinen auf dem Markt im Jahre 1979 sehr schnell Verbreitung gefunden, zumal auch Probleme, die vorher mit dem 8Bit-Mikroprozessor gelöst wurden, mit ihm oft wesentlich einfacher und bequemer bewältigt werden können. Der 16Bit-Mikroprozessor scheint ein an den Schwierigkeitsgrad sehr vieler Informationsverarbeitungsprobleme in der Praxis besonders gut angepaßter integrierter Schaltkreise zu sein. Trotzdem wird er den 8Bit-Mikroprozessor wohl nicht aus seiner Vorrangstellung verdrängen, und auch das Vordringen anderer Varianten, insbesondere von 1Bit- und 32Bit-Mikroprozessoren, nicht verhindern. Der 4Bit-Mikroprozessor scheint dagegen tatsächlich zunehmend an Bedeutung zu verlieren. Noch vor kurzer Zeit entfielen grob gerechnet 90% des Umsatzes auf den 8Bit- und 10% auf den 16Bit-Mikroprozessor. In der Einleitung haben wir hervorgehoben, daß die Mikroelektronik ein sehr komplexes Gebiet ist, das

unter verschiedenen Aspekten behandelt werden kann. Eine Reihe dieser Aspekte — die wichtigsten, wie uns für ein erstes Vertrautwerden mit diesem Gebiet scheint — haben wir in den vorangegangenen Kapiteln behandelt. Einer ist übriggeblieben, oder wurde bisher doch nur am Rande gestreift, nämlich die künftige Entwicklung der Mikroelektronik. Darauf soll im letzten Kapitel eingegangen werden.

7. Perspektiven der Mikroelektronik

Ebenso wie die heutige Mikroelektronik aus dem Bedürfnis nach einer besseren Technik für die Informationsverarbeitung hervorgegangen ist, entstehen von daher auch die Antriebe für ihre weitere Entwicklung. Mikroelektronik ist nicht identisch mit Informationsverarbeitungstechnik. Sie ist ihre materielle Grundlage, das Baumaterial gewissermaßen, aus dem diese Technik ihre Gebäude errichtet. Den Bauplan liefert die Mikroelektronik selbst nicht, dafür sind andere Zweige der Informationsverarbeitungstechnik, die der Mathematik nahestehen und die man oft unter dem Begriff «Computerwissenschaft» oder «Informatik» zusammenfaßt, zuständig. Fortschritte der Informationsverarbeitungstechnik machen nicht allein Fortschritte der Mikroelektronik, sondern auch Fortschritte der Informatik erforderlich. Als z. B. Mitte der siebziger Jahre die Computer eine solche Leistungsfähigkeit erreicht hatten, daß Programme mit vielen hunderttausend Befehlen bearbeitet werden konnten, traten prinzipielle Schwierigkeiten bei der Programmierung auf. Die damals auch vom besten Programmierer nicht wesentlich unterschreitbare Grenze von fünf Fehlern pro tausend Programmschritte, also von einigen hundert Fehlern in einem Programm aus hunderttausend Schritten, war

viel zu groß, um durch Tests gefunden und ausgemerzt werden zu können. Man sprach von einer Software-Krise. Die Informatik hilft, diese Krise zu überwinden. Z. B. hat sie Regeln für die Disziplinierung des Programmierens entwickelt, die die Fehlerrate auf ein Zehntel des früheren Wertes zu reduzieren gestatteten. Heute, da es in der Informationsverarbeitungstechnik u. a. um den Bau noch leistungsfähigerer großer Rechner geht, ist auch die Frage nach dem Aufbauprinzip, der Architektur von Großrechnern, neu aufgeworfen. Die «von-Neumann-Architektur», nach der heute nahezu alle Rechner funktionieren, ist ins Gerede gekommen. Von der Computerwissenschaft werden neue Lösungen für die Rechnerarchitektur erwartet. Sie versucht dazu u. a., von dem konsequenten zeitlichen Nacheinander der einzelnen Arbeitsschritte beim von-Neumann-Rechner abzugehen und da, wo es möglich ist, die zeitlich parallele Durchführung mehrerer Schritte anzuwenden. Erste Ergebnisse auf diesem Wege wurden bereits erzielt. Im Zusammenhang mit einer neuen Rechnerarchitektur wird aber auch über die Nutzung biologischer Prinzipien der Evolution und Selbstorganisation spekuliert.

Wir haben diese Bemerkungen vorangestellt, um den folgenden Aus-

führungen, die sich nur mit der Mikroelektronik und den von *ihr* zu leistenden Beiträgen für den Fortschritt der Informationsverarbeitungstechnik beschäftigen werden, den richtigen Rahmen zu geben. Die integrierten Schaltkreise, wie sie heute von der Mikroelektronik-Industrie hergestellt werden, leisten bereits außerordentlich viel. Sie sind billig, sie lösen anspruchsvolle Informationsprobleme zuverlässig, schnell und energiesparend und sind in großer Breite einsetzbar. Aber das ist nur die halbe Wahrheit. Zuverlässig, schnell, sparsam, anspruchsvoll, breit oder billig sind relative Begriffe. Mißt man die heute verfügbaren LSI- und VLSI-Schaltkreise an den elektronischen Hilfsmitteln zur Informationsverarbeitung aus der Vergangenheit, ganz zu schweigen von den mechanischen bzw. elektromechanischen, so haben diese Begriffe ihre volle Berechtigung. Orientiert man sich aber an den Erfordernissen, wie sie sich aus den in der Praxis zu lösenden Informationsproblemen ergeben, so kommt man zu einer anderen Einschätzung. Es geht dabei weniger um den Platzbedarf, den Energieverbrauch und die Zuverlässigkeit der Schaltkreise. Diese Eigenschaften haben ein Niveau erreicht, mit dem man vorerst zufrieden sein könnte, obwohl natürlich auch hier weitere Verbesserun-

gen wünschenswert sind. Aber dafür sind nicht unbedingt neuartige Schaltkreise oder prinzipiell andere Herstellungstechnologien erforderlich. Vieles läßt sich hier auch schon durch bessere Ausschöpfung bereits eingeführter Technologien erreichen. An dieser Stelle ist es instruktiv, die Situation der Mikroelektronik mit der der Autoindustrie zu vergleichen. Die heute hergestellten Autos sind zwar weder ausreichend zuverlässig noch billig, noch ist ihr Treibstoffverbrauch gering genug. Aber sie sind vom Preis her erschwinglich und erfüllen ihren Zweck insgesamt schon recht gut, insbesondere sind sie ausreichend schnell. Um die wünschenswerten Verbesserungen zu erreichen, muß sich die Autoindustrie nicht nach einem völlig neuen Transportmittel umsehen, sondern kann versuchen, diesem Ziel durch die Entwicklung besserer Autotypen und die Verbesserung der Technologie ihrer Herstellung näherzukommen. Im Hinblick auf den Platzbedarf, den Energieverbrauch und die Zuverlässigkeit ist die Situation bei den integrierten Schaltkreisen ähnlich. Wesentliche Unterschiede bestehen aber in anderer Hinsicht.

Intelligenterer

Informationsverarbeitungssysteme

Durch das Auto ist heute das Problem der Fortbewegung entlang von Straßen, man darf wohl sagen, vollständig und endgültig gelöst. Mit den heute erhältlichen integrierten Schaltkreisen und den aus ihnen herstellbaren Informationsverarbeitungssystemen läßt sich aber nur ein Teil der anstehenden Probleme überhaupt lösen, und bei weitem nicht alle Lösungen sind so effektiv wie man das wünschen muß. Man

vergegenwärtige sich dazu nur einmal, welche komplizierten Probleme durch den Menschen gelöst werden können bzw. heute noch durch ihn gelöst werden müssen, weil die mikroelektronischen Schaltkreise und die aus ihnen aufgebauten Systeme für die Informationsverarbeitung nicht intelligent genug dafür sind. Sprachübersetzungen können heute selbst von den besten Rechnern nur mit einem eingeschränkten Wortschatz und nur mit großen stilistischen Unzulänglichkeiten ausgeführt werden. Die leistungsfähigsten elektronischen Rechner sind heute noch nicht in der Lage, einen Großmeister im Schach zu schlagen. Probleme können heutige Rechner überhaupt nur dann lösen, wenn ihnen der Lösungsweg bis zu einem bestimmten Grade vorgegeben wird und wenn das außerdem in einer Programmiersprache geschieht, die vom Rechner verstanden werden kann, die sich aber von natürlicher Sprache noch sehr stark unterscheidet. Es besteht kein Zweifel, die Fähigkeit von Computern zur Lösung anspruchsvoller Probleme der Informationsverarbeitung, ihre künstliche Intelligenz, steht heute noch weit hinter der natürlichen Intelligenz des Menschen zurück. Nicht zuletzt deshalb ist der Bau immer intelligenterer Computer eine aktuelle Aufgabe für die Informationstechnik. Da der Intelligenzgrad eines Computers, wenn auch nicht ausschließlich, so doch wesentlich durch die Anzahl, die Schnelligkeit und den Integrationsgrad der zu seinem Aufbau verwendeten Schaltkreise bestimmt ist — sehr viele, sehr schnelle und hochintegrierte Schaltkreise ermöglichen eine hohe Intelligenz —

ergeben sich daraus unmittelbar Anforderungen an die Mikroelektronik. Seit einigen Jahren wird in Japan und in der Sowjetunion an der Schaffung von Computern eines bisher nicht gekannten Intelligenzgrades gearbeitet. Man nennt sie Rechner der fünften Generation. Ein solcher Rechner soll Aufgaben verstehen, die in natürlicher Sprache geschrieben oder gesprochen sind. Fehler in der Aufgabenstellung sollen von ihm erkannt und selbstständig korrigiert werden. Stößt der Rechner bei der Lösung einer Aufgabe auf neue Informationen, Regeln oder Gesetzmäßigkeiten, so sollen sie von ihm registriert und in seinem Speicher festgehalten werden. Der Rechner soll nicht nur Daten, sondern auch *Wissen* verarbeiten. Abschätzungen haben ergeben, daß ein solcher Rechner in der Lage sein müßte, pro Sekunde etwa 10^{11} Befehle auszuführen und Arbeitsspeicher mit einer Kapazität von vielen MByte besitzen müßte. Daraus entsteht an die Mikroelektronik die Forderung, Mikroprozessoren mit Befehlsabarbeitungszeiten im Nanosekundenbereich und mit größeren Verarbeitungsbreiten als bisher, d. h. mit 32 Bit und darüber, sowie Halbleiterspeicher zur Verfügung zu stellen, die kürzere Zugriffszeiten und Speicherkapazitäten deutlich über 256 KBit bis hin zu mehreren MBit besitzen. Für die japanischen Rechner der fünften Generation sind integrierte Schaltkreise zunächst mit 10^6 , später mit 10^7 Bauelementen vorgesehen. Ohne Schaltkreise mit derart hohen, heute noch nicht beherrschten Integrationsgraden oder mit völlig neuartigen Funktionselementen, auf die wir noch eingehen werden, sind die

Rechner der fünften Generation nicht denkbar.

Höchstmögliche maschinelle Intelligenz zu einem niedrigen Preis

Sollte das japanische Konzept aufgehen — nicht wenige zweifeln allerdings daran — wären die Rechner natürlich sehr teuer. Aber wenn es primär um höchstmögliche maschinelle Intelligenz geht, spielen die Kosten zunächst nur eine untergeordnete Rolle. Bei weniger exklusiven Rechnern, die für eine Vielzahl von Anwendungen in hoher Stückzahl benötigt werden, kommt es allerdings auf den Preis entscheidend an. Hier ist möglichst hohe maschinelle Intelligenz für einen möglichst niedrigen Preis gefragt. Würde die gleiche Intelligenz auf der Basis mikroelektronischer Schaltkreise heute für ein Zehntel des Preises auf dem Weltmarkt zu haben sein, den sie tatsächlich kostet, könnte eine Reihe von Informationsverarbeitungsproblemen in allen Bereichen der Praxis auf einer qualitativ höheren Stufe gelöst werden als das zur Zeit der Fall ist. Wenn heute diese höhere Stufe international noch nicht erreicht wird, so liegt das nicht daran, daß die erforderliche maschinelle Intelligenz auf dem Markt prinzipiell noch nicht zu haben ist, sondern daran, daß sie zu teuer ist. Nehmen wir z. B. die automatische Steuerung von Prozessen in der Industrie. Dabei liegt die Information zunächst in analoger Form vor, muß dann in die digitale Form umgewandelt, so verarbeitet und anschließend wieder in die analoge Form rückverwandelt werden. All das muß so schnell geschehen, daß der weitere Ablauf des Prozesses noch rechtzeitig beeinflußt werden kann. Man spricht in

diesem Zusammenhang von On-line-Verarbeitung und meint damit, daß Informationsverarbeitung und realer Prozeß gleichzeitig miteinander, sozusagen in einer Linie, laufen. Je schneller der Prozeß ist und je mehr Information über ihn verarbeitet werden soll, um so schneller muß das Informationsverarbeitungssystem sein. Die Schnelligkeit der Informationsverarbeitung besitzt hier eine absolutere Bedeutung als etwa bei einem Rechner zur Lösung mathematischer Probleme. Es muß eine bestimmte Mindestgeschwindigkeit überschritten werden, damit überhaupt eine On-line-Verarbeitung möglich wird. Erfüllt das Informationsverarbeitungssystem diese Forderungen nicht, d. h., ist seine maschinelle Intelligenz zu gering oder ist das System, das die erforderliche Intelligenz besitzt, zu teuer, um eingesetzt werden zu können, muß das Niveau der Informationsverarbeitung und damit der Prozeßsteuerung niedriger angesetzt werden, als es eigentlich wünschenswert wäre. Tatsächlich ist heute die für einen tolerierbaren Preis zu habende maschinelle Intelligenz in vielen Fällen noch zu gering, um technische Prozesse, insbesondere schnell ablaufende, bestmöglich steuern zu können. Das trifft z. B. auf Industrieroboter der dritten Generation zu, die mit einer möglichst umfangreichen Sensorik für die Erfassung des Zustandes des Bearbeitungsobjektes und der Umwelt ausgerüstet werden sollen. Dabei macht vor allem die Verarbeitung von Bildern einen hohen Aufwand, weil Bilder in der Regel einen hohen Informationsgehalt haben. Ganz allgemein erfordert die schnelle digitale Verarbei-

tung analoger Signale Informationsverarbeitungssysteme mit einem relativ hohen Grad an maschineller Intelligenz. Die Ursachen dafür haben wir in Kapitel 4 genannt. Trotzdem gibt es Gründe, von der analogen Informationsverarbeitung zur digitalen überzugehen. Wir sind darauf ebenfalls bereits eingegangen. Wenn erst höher integrierte, schnellere Schaltkreise, darunter schnellere AD-Wandler zur Verfügung stehen werden, wird man viele Informationsverarbeitungsprobleme, für die heute noch die analoge Verarbeitungstechnik eingesetzt wird, in digitaler Form lösen. Dazu gehören nicht zuletzt auch Probleme aus dem großen Gebiet der Unterhaltungselektronik. Für die digitale Verarbeitung der analogen Bildsignale in einem Farbfernsehgerät zum Beispiel ist ein AD-Wandler, der pro Sekunde 10^7 analoge Signale in jeweils 8 Bit breite Datenwörter umwandelt und ein hochintelligenter Speziialschaltkreis für die schnelle digitale Informationsverarbeitung erforderlich. Wollte man einen Mikroprozessor für die Verarbeitung einsetzen — was prinzipiell möglich, praktisch aber höchst unzumutbar wäre — so müßte dieser bei 8 Bit Verarbeitungsbreite eine Taktfrequenz von etwa 100 MHz zulassen. Erste teilweise digital arbeitende Fernsehgeräte sind bereits öffentlich vorgestellt worden. Sie haben ein schärferes Bild als die Analoggeräte, das Flimmern entfällt bei ihnen. Auch digitale Rundfunkempfänger mit High-Fidelity-Qualität sind im Kommen. Am weitesten fortgeschritten ist die Nutzung der Digitaltechnik bei den Plattenspielern. Das zu verarbeitende Signal — von der

Platte mittels Laserstrahl abgetastete Hell- und Dunkelwerte — ist dabei im Unterschied zum Fernsehen und Rundfunk von vornherein digital. Die ersten digital arbeitenden Plattenspieler sind bereits auf dem Markt. Sie werden gewiß nicht die letzten Digital-Geräte der Unterhaltungselektronik bleiben. Ein Übergang zur Digitaltechnik vollzieht sich heute auch bei der Informationsübertragung. Ein digitales Signal, das nur 0 oder 1 sein kann, läßt sich wesentlich sicherer wiedererkennen, nachdem es einen Übertragungskanal mit seinen unvermeidlichen Störungen, dem sogenannten Rauschen, durchlaufen hat, als ein analoges mit seinen vielen kontinuierlichen Abstufungen. Telefoniert wird heute nur noch zum Teil durch die Übertragung analoger Signale über Koaxialkabel aus Kupfer. In den Telefonnetzen vieler Länder vollzieht sich gegenwärtig der Übergang zur digitalen Technik. Wird als Übertragungsmedium nicht ein Kupferkabel, sondern ein Lichtleiter verwendet, kommt praktisch nur noch die Digitaltechnik in Frage. Durch den Lichtleiter lassen sich bei einer Wellenlänge von $1,3\text{ }\mu\text{m}$ theoretisch über 10^{14} Bit/s übertragen. Um diese ungeheure Übertragungskapazität möglichst weitgehend ausnutzen zu können, benötigt man schnelle elektronische Schaltkreise, die die Information digital codiert in den Sender (eine Lumineszenz- oder Laserdiode) einspeisen und vom Empfänger (einem Fotodetektor) wieder abnehmen können. Auch für die drahtlose Telekommunikation über gerichtet oder ungerichtet abgestrahlte elektromagnetische Wellen ist die Digitaltechnik von Bedeu-

tung. Bei der Nachrichtenübertragung über Richtfunkstrecken und über Nachrichtensatelliten mit Hilfe von Mikrowellen wird sie zum Teil heute schon eingesetzt. Aber auch für die gegenwärtig noch dominierende analoge Mikrowellen-Übertragungstechnik sind schnellere Schaltkreise erforderlich, als sie heute zur Verfügung stehen. Man muß deshalb in der Mikrowellen- oder Höchstfrequenztechnik heute auch noch schnelle Elektronenröhren — z. B. Magnetrons oder Klystrons — einsetzen. Das macht diese Technik sehr aufwendig und teuer und steht der wünschenswerten Ausweitung ihrer praktischen Nutzung — man denke z. B. an den Satelliten-Direkttempfang — als Hindernis im Wege.

Höhere Integrationsgrade

und höhere Arbeitsgeschwindigkeit

Die bestmögliche Steuerung technischer Prozesse, der Übergang von der analogen zur digitalen Informationsverarbeitung und -übertragung, die Einführung der Lichtleitertechnik oder der breitere Einsatz und weitere Ausbau der Mikrowellenübertragung sind nur einige Beispiele für Entwicklungen in der Informationstechnik, die Informationsverarbeitungssysteme mit einer höheren Intelligenz für einen niedrigeren Preis erforderlich machen. Höhere Intelligenz wird möglich durch mehr und schnellere Bauelemente, oder auch *nur* durch schnellere Bauelemente. Die Erhöhung der Bauelementezahl läßt sich rein arithmetisch auch dadurch erreichen, daß mehr Schaltkreise mit unverändert niedrigem Integrationsgrad eingesetzt werden. Aber diese Lösung ist teuer, in vielen Fällen zu teuer. Es

ist die Lösung, die wir weiter oben als im Prinzip heute bereits erhältlich, aber aus Kostengründen praktisch nicht einsetzbar charakterisiert hatten. Es kommt nur eine Lösung in Frage, bei der die Vergrößerung der Zahl der eingesetzten Bauelemente mit einer Verringerung des Preises eines einzelnen Bauelements einhergeht. Eine solche Lösung ist der Einsatz höher integrierter Schaltkreise. Wir erinnern in diesem Zusammenhang an die Ausführungen in Kapitel 4, die deutlich gemacht haben, warum und unter welchen Voraussetzungen die Kosten eines Bauelements durch Übergang zu höher integrierten Schaltkreisen gesenkt werden können. Von dieser Möglichkeit kann und muß die Mikroelektronik Gebrauch machen, wenn sie höhere Intelligenz zu einem niedrigeren Preis zur Verfügung stellen soll. Noch in einem anderen Punkt ist die Lösung mit einem höher integrierten Schaltkreis der mit mehreren niedriger integrierten Schaltkreisen vorzuziehen: Sie ist zuverlässiger. Die leitenden Verbindungen, die zwischen den verschiedenen Schaltkreisen bei der Variante mit niedrigerem Integrationsgrad hergestellt werden müssen, sind im allgemeinen relativ störanfällig, jedenfalls verglichen mit den Verbindungen, die bereits bei der Fertigung eines integrierten Schaltkreises im Zyklus 1 zwischen den einzelnen Bauelementen hergestellt werden. Wünscht man eine hohe Zuverlässigkeit, so muß man die Zahl der nachträglich anzubringenden Verbindungen möglichst klein halten, d. h. mit möglichst hoch integrierten Schaltkreisen arbeiten.

Wir fassen zusammen: Nicht nur

die Forderung nach Informationssystemen mit einer möglichst hohen Intelligenz, von der am Anfang dieses Kapitels die Rede war, sondern auch die Forderungen nach Systemen mit einem günstigen Verhältnis von Intelligenzgrad und Preis und einer höheren Zuverlässigkeit zwingen zur Höherintegration. Der Mikrorechner aus wenigen hochintegrierten Schaltkreisen ist ein Produkt dieses Zwanges. Aber schon geht die Entwicklung zu Mikro- und Minirechnern auf einem einzigen Chip hin. Die ersten Typen sind bereits auf dem Markt. Allerdings erreichen diese Einchipmikro- und Minirechner heute noch nicht die Leistungsfähigkeit ihrer aus mehreren Schaltkreisen zusammengesetzten Artgenossen. Sollen größere als Mikro- und Minirechner aus möglichst wenigen Schaltkreisen aufgebaut werden, benötigt man Mikroprozessoren mit 32 Bit Verarbeitungsbreite und darüber und Speicherschaltkreise mit Kapazitäten über 64 KBit. Einen solchen Rechner hat z. B. die Firma Hewlett Packard unter der Typenbezeichnung HP9000 entwickelt. Der 32Bit-Mikroprozessor für diesen Rechner enthält 450 000 Bauelemente auf einem Chip. Ein für den Rechner bestimmter Speicherbaustein, ein 128 KBit RAM, verfügt sogar über noch mehr, nämlich 600 000 Bauelemente. Daß gerade der Speicherschaltkreis des Rechners den höchsten Integrationsgrad erreicht, ist typisch. Der Grund dafür ist die relativ einfache Struktur der Speicherschaltkreise: ein und dasselbe Grundelement, die Speicherzelle für 1 Bit, wird vielfach wiederholt. Speicherschaltkreise mit Kapazitäten bei

oder über 1 MBit würden, wenn sie zu einem niedrigen Preis zu Verfügung stünden, übrigens nicht nur für Rechner, sondern auch für viele andere Zwecke einsetzbar sein. Zum Beispiel könnte man in einem 1 MBit ROM ein Buch mit etwa 100 Seiten speichern. Zunehmend höhere Integrationsgrade aufgrund wachsender Anforderungen erreichen auch AD- und DA-Wandler. Nehmen wir z. B. einen 10 Bit/100 MHz-AD-Wandler. Würde er nach dem in Kapitel 4 beschriebenen Prinzip der Parallelumsetzung, das die Befriedigung hoher Geschwindigkeitsanforderungen gewährleistet, realisiert werden, so würde man $2^{10} - 1 = 1023$ Komparatoren benötigen. Rechnet man für einen Komparator 100 Bauelemente, so kommt man auf mehr als 100 000 Bauelemente für den AD-Wandler, wobei die für die digitale Codierung erforderlichen Bauelemente noch gar nicht mitgezählt sind. Besonders hoch sind die Anforderungen an die Arbeitsgeschwindigkeit der AD-Wandler. Solen, wie in unserem Beispiel, pro Sekunde 10^8 Analogsignale digitalisiert werden, muß die Verzögerungszeit eines Logik-Gatters im Schaltkreis deutlich unter 10^{-8} Sekunden liegen. Die Erhöhung der Arbeitsgeschwindigkeit von Schaltkreisen bedarf noch einer weiteren Erörterung. Während Intelligenzgrad und Preis pro Bauelement bei einem integrierten Schaltkreis direkt mit dem Integrationsgrad verbunden sind — die Intelligenz steigt, der Preis pro Bauelement fällt unter bestimmten Voraussetzungen, wenn der Integrationsgrad erhöht wird — trifft das auf die Arbeitsgeschwindigkeit nicht im

gleichen Maße zu. Integrationsgrad und Arbeitsgeschwindigkeit sind in gewissen Grenzen voneinander unabhängig. Man kann höher integrieren, ohne daß dadurch der Schaltkreis schneller wird. Die Erhöhung der Arbeitsgeschwindigkeit muß man also auch direkt angehen. Die Wege dazu sind in den vorangegangenen Kapiteln zum Teil schon angedeutet worden. Es sind dies der Ersatz von Silizium durch Materialien mit höheren Ladungsträgerbeweglichkeiten, insbesondere die Verwendung von Galliumarsenid und Indiumphosphid, der Übergang zu niedrigeren Betriebstemperaturen und die Verkleinerung der Bauelemente. Von kleineren Bauelementen lassen sich aber mehr in einem Schaltkreis unterbringen als von größeren. Dadurch ergibt sich dann doch wieder ein gewisser Zusammenhang zwischen der Arbeitsgeschwindigkeit und dem Integrationsgrad, wenn auch ein weniger direkter als bei der Intelligenz und dem Bauelemente-Preis. Nachdem wir uns über die Gründe Rechenschaft abgelegt haben, die die Mikroelektronik zur Höherintegration treiben, wollen wir uns die Wege etwas genauer ansehen, die zur Erreichung dieses Ziels beschritten werden. Theoretisch gibt es dafür wenigstens fünf: Erstens die Verkleinerung der Bauelementeabmessungen, zweitens die Vergrößerung der Chipfläche, drittens die Integration in mehreren Ebenen, also die Einbeziehung der dritten Raumdimension in die Integration, viertens die Komprimierung von mehreren Funktionen in einem einzigen Bauelement, wie das z. B. beim Bipolar-Transistor mit mehreren Emittoren

der Fall ist, und fünftens schließlich die Entwicklung platzsparenderer Schaltungen und Realisierungstechniken. Bei den zwei zuletzt genannten Punkten wollen wir es mit der Erwähnung bewenden lassen. Auf die dreidimensionale Integration werden wir später noch kurz eingehen. Die Vergrößerung der Chipfläche — angefangen hat man mit weniger als 1 mm^2 im Jahre 1960, heute sind einige . . . zig Quadratmillimeter erreicht — stellt sicher eine wichtige Reserve für die Höherintegration dar. Wie wir in Kapitel 5 bereits erwähnt haben, wird aber die Ausbeute mit wachsender Chipfläche kleiner, es sei denn, man ist in der Lage, Substratscheiben mit einer geringeren Anzahl von Fehlern pro cm^2 herzustellen. Nur in dem Maße, in dem das gelingt, wird man in Zukunft zu Chipflächen bei oder sogar über 1 cm^2 übergehen können. Das würde dann aber auch mindestens eine Verzehnfachung des Integrationsgrades bedeuten. Der Hauptweg zur höheren Integration ist **Submikrometer-Technologie** heute aber zweifellos die weitere Verkleinerung der Bauelementeabmessungen. 1960 waren die kleinsten Strukturen, die man erzeugen konnte, etwa 30 Mikrometer breit. Heute werden von der Industrie Spitzenwerte erreicht, die zwischen 2 Mikrometern und 1 Mikrometer liegen. Sie werden bei den größten integrierten Schaltkreisen realisiert, die bisher produziert wurden, etwa einem 256KBit-RAM oder einem 32Bit-Mikroprozessor mit jeweils einigen 10^5 Bauelementen. Um Integrationsgrade von 10^6 und 10^7 zu erreichen, muß man Strukturen kleiner als 1 Mikrometer beherrschen. Das ist

eine der entscheidendsten Aufgaben für die weitere Entwicklung der Mikroelektronik. Ihre Lösung macht Veränderungen der gesamten Technologie, wie wir sie in Kapitel 5 beschrieben haben, erforderlich. Die einzelnen technologischen Schritte müssen zum Teil bis an ihre prinzipiellen Grenzen hin ausgenutzt werden, zum Teil werden diese Grenzen aber bereits überschritten, und es müssen neue Verfahren eingeführt werden.

Die Grenze der *Fotolithographie* ist durch die Wellenlänge des Lichtes bestimmt. Verwendet man statt Licht Röntgenstrahlen oder Elektronenstrahlen, so lassen sich Strukturen bis zu hundertsteln von Mikrometern erreichen. An der Entwicklung der Röntgen- und Elektronenstrahlolithographie wird heute ebenso gearbeitet wie an der vollen Ausschöpfung der Fotolithographie, die ebenfalls Strukturierungen im Submikrometerbereich zulassen sollte. Für die Röntgenstrahlolithographie ist das Elektronen-Synchrotron, das in der Kernphysik entwickelt wurde, als Strahlungsquelle interessant. Für die Fotolithographie mit UV-Strahlung gewinnt der Excimer-Laser Bedeutung. Einen ganz anderen Weg zur Verkleinerung der Strukturabmessungen beschreitet man bereits seit längerem mit der sogenannten VMOS-Technologie. Das «V» steht für «Vertikal» und bedeutet, daß der normalerweise horizontal ausgehende Kanal eines MOS-Transistors vertikal im Chip untergebracht ist. Da die vertikalen Strukturabmessungen durch die Dicken der verschiedenen Materialschichten bestimmt sind und diese bei oder sogar unter $1 \mu\text{m}$ liegen, lassen sich auf

diese Weise sehr kleine Kanallängen erzielen. Beim bipolaren Transistor ist die zur Kanallänge des FET analoge kritische Abmessung, die Basisbreite, ohnehin nicht durch die lateralen Strukturierungsverfahren bestimmt, sondern durch die vertikale Ausdehnung und Schärfe des Dotierungsprofils, das zur Erzeugung der Basis hergestellt werden muß. Für beide Arten von Bauelementen gilt, daß mit kleiner werdenden horizontalen Abmessungen auch die vertikalen Ausdehnungen, also die Dicken der Bauelementeschichten kleiner gemacht werden können bzw. müssen. Aus der Verkleinerung der Bauelementestrukturen resultieren höhere Anforderungen an die gesamte Halbleitertechnologie, d. h. insbesondere an die Verfahren zum Dotieren, Oxidieren, Ätzen, Reinigen, Passivieren und Metallisieren. Die *Ionenimplantation* mit anschließender Materialausheilung gewinnt als Dotierungsverfahren gegenüber der Diffusion an Bedeutung, nicht zuletzt deshalb, weil mit kleiner werdenden vertikalen Strukturen die relativ geringen Dotierungstiefen der Ionenimplantation ausreichen. Im Vergleich zur Diffusion ermöglicht die Ionenimplantation schärfere Dotierungsprofile. Vor allem entfallen aber die massiven thermischen Belastungen durch längeres Erhitzen der Siliziumscheiben über 1000°C bei der Diffusion. Wird die Implantation mit einem fein gebündelten Ionenstrahl ausgeführt, können die erforderlichen Strukturen ohne Zuhilfenahme der Fotolithographie durch geeignete Führung des Strahls direkt erzeugt werden. Aus der Vermeidung länger anhaltender hoher Temperaturen bei Anwendung der Ionenimplantation

ergibt sich allerdings ein Problem für die Erzeugung der Siliziumdioxidschicht vor jedem Dotierungsschritt. Soll die thermische Oxidation auch bei den dann unter Umständen wesentlich niedrigeren Temperaturen genügend schnell ablaufen, muß man den Sauerstoff unter hohen Druck setzen — oder gänzlich auf die thermische Oxidation verzichten und ein anderes Verfahren für die Oxidbildung einsetzen. Andere Probleme bietet das *Reinigen* der Substratscheiben mit Wasser und das *Ätzen* derselben mit flüssigen Chemikalien. Diese Prozesse waren schon immer gewisse Fremdkörper innerhalb der Halbleitertechnologie. Beim Übergang zu Strukturen im Mikro- und Submikrometerbereich werden sie vollends unhaltbar und müssen durch besser kontrollierbare Verfahren wie das Reinigen bzw. Ätzen in einem Gasplasma oder mit Hilfe von Ionenstrahlen ersetzt werden. Beide Verfahren ermöglichen auch das direkte Herausätzen von definierten Strukturen ohne Zuhilfenahme der Fotolithographie. Beim Plasmaätzen z. B. kann man durch bestimmte Maßnahmen erreichen, daß nur diejenigen Stellen auf dem Substrat weggeätzt werden, die zusätzlich der UV-Strahlung eines Excimer-Lasers ausgesetzt sind. Zur Passivierung, Maskierung und zur Isolation werden neben dem Siliziumdioxid zunehmend auch andere Materialien, insbesondere Siliziumnitrid, mit zum Teil besseren Eigenschaften, als sie das Siliziumdioxid aufweist, verwendet. Diese und andere Schichten werden in wachsendem Maße durch Verfahren erzeugt, die mit relativ niedrigen Temperaturen auskommen, wie z. B. die CVD

(Chemical Vapour Deposition = chemische Abscheidung aus der Dampfphase), die durch eine elektrische Plasmaentladung oder die Strahlung eines Lasers noch zusätzlich unterstützt werden kann. Mit kleiner werdenden horizontalen Strukturen der Bauelemente werden auch die metallischen Verbindungen zwischen ihnen schmaler. Die auftretenden Stromdichten nehmen extrem hohe Werte an. Dadurch kann es zu Materialzerstörungen, der sogenannten Elektromigration, kommen. Es müssen Methoden gefunden werden, widerstandsfähigere Leitbahnen aus Aluminium oder anderen Metallen herzustellen. Dabei gewinnen Silizide des Wolfram oder Molybdän zunehmend an Bedeutung. Die Silizide eignen sich darüber hinaus auch für Gate- und Kontaktelektroden.

Diese Beispiele für Entwicklungen in der Halbleitertechnologie beim Übergang zur Submikrometer-Elektronik mögen genügen. Allgemein kann man sagen, daß dieser Übergang die Eliminierung von Hochtemperatur- und naßchemischen Schritten aus der Halbleitertechnologie und ihre Ersetzung durch Niedertemperatur- und trockene Verfahren erfordert. Um die notwendige hohe Präzision der Verfahrensschritte gewährleisten zu können, müssen dieselben diagnostisch verfolgt und über Mikrorechner automatisch gesteuert werden. Der Einsatz moderner Elektronen-, Ionen-, Röntgen- und optischer Diagnoseverfahren erhält in diesem Zusammenhang eine besondere Bedeutung. Gleichzeitig erhöhen sich beim Übergang zur Submikrometer-Technologie die auch heute schon extremen

Reinheitsanforderungen an Materialien und Verfahren noch einmal ganz erheblich. Während es z. B. bei der «Über-1 Mikrometer-Technologie» ausreicht; in den Reinräumen darauf zu achten, daß die Zahl der Staubteilchen mit einer Größe über $0,5\text{ }\mu\text{m}$ einen bestimmten Grenzwert nicht überschreitet, muß man bei der Submikrometer-Technologie Staubteilchen bis herab zu einem Zehntel Mikrometer Größe aus den Reinräumen weitgehend entfernen. Symbolisch und etwas zugespitzt könnte man sagen, daß die Reinräume, die bisher nur frei von Bakterien zu sein brauchten — das Tuberkelbakterium ist $1\text{ }\mu\text{m}$ lang — in Zukunft auch frei von Viren sein müssen — die Größe des Pockenvirus beträgt $0,15\text{ }\mu\text{m}$. Diese und die anderen, vorher genannten Anforderungen an eine künftige Submikrometer-Technologie könnten aber erst dann richtig befriedigt werden, wenn man zum vollautomatisierten Zyklus 1 im hermetisch abgeschlossenen Volumen, möglichst im Vakuum, überginge. Dadurch würden dann auch alle Beeinträchtigungen des Herstellungsprozesses durch das Bedienungspersonal, wie z. B. das Verursachen von Erschütterungen oder Verunreinigungen, entfallen. Bedienungsfehler wären praktisch ausgeschlossen. Experten schätzen ein, daß allein durch die Vermeidung von Bedienungsfehlern die Ausbeute im Zyklus 1 auf das Drei- bis Vierfache gesteigert werden könnte. Der durchgehend automatisierte Vakuumprozeß im Zyklus 1 wird aber, wenn überhaupt, nur schrittweise zu erreichen sein. Real ist zunächst die Automatisierung von Teilschritten, etwa der Fotolithografie.

Submikrometer-Bauelemente

Schaltkreise mit Integrationsgraden von 10^6 und 10^7 und vielleicht sogar noch darüber und, was unausweichlich ist, mit Strukturen deutlich unter $1\text{ }\mu\text{m}$ sind also eine starke Herausforderung an die Mikroelektronik. Es sind entscheidende Verbesserungen der gesamten Herstellungstechnologie im Zyklus I erforderlich. Dazu kommen aber auch Probleme, die mit der Funktion der Bauelemente selbst zusammenhängen. Kann man annehmen, daß ein MOS-FET unverändert gut funktioniert, wenn man ihn immer kleiner macht? Mit Sicherheit nicht. Er muß z. B. mindestens noch ein Siliziumatom enthalten. Aber das ist bei weitem nicht ausreichend. Es muß p- und n-leitendes Material vorhanden sein, und das bedeutet, daß die p- und n-Gebiete mindestens ein Dotierungsatom enthalten müssen. Bei einer Dotierungskonzentration von 10^{18} cm^{-3} ist dafür im Mittel ein Volumen von 10^{-18} cm^3 erforderlich. Das ist ein Würfel mit einer Kantenlänge von $1/100\text{ }\mu\text{m}$. Aber ein Dotierungsatom ist natürlich noch viel zu wenig, in Wirklichkeit müssen es viele sein. Damit sich ein p-Gebiet zwischen Source und Drain überhaupt ausbilden kann, muß der Kanal eine bestimmte Länge (ein Mehrfaches der Debye-Länge) überschreiten, sonst kommt es zum «Durchgreifen» (punch through). Auch damit der Ladungsträgertransport im Kanal «normal» erfolgt, ist eine Mindestlänge (die mittlere freie Weglänge) notwendig, sonst wird der Transport «Ballistisch». Ist die Wellenlänge der Elektronen bei einem n-MOS-FET nicht mehr klein gegenüber der Tiefe des Kanals, kommt

es zu Quantisierungseffekten — die Elektronen im Kanal werden zu einem 2-dimensionalen Elektronengas. Um den MOS-FET noch sicher aus dem gesperrten in den leitenden Zustand schalten zu können, muß die Spannungsänderung am Gate sehr viel größer als die thermischen Spannungsschwankungen sein. Bei Zimmertemperatur betragen diese etwa ein vierzigstel Volt. Die gegenseitige Isolation der Bauelemente durch pn-Übergänge, die in Sperrrichtung gepolt sind, funktioniert nur solange wie die Energielücke nicht durchtunnelt werden kann, d. h., solange die Dicke der Raumladungsschicht noch groß gegenüber der Wellenlänge der Elektronen oder Löcher ist. Heute wird beim Entwurf von hochintegrierten Schaltkreisen weitgehend vom sogenannten «scaling down», dem «Herabskalieren» der Bauelemente Gebrauch gemacht. Das «scaling down» ermöglicht Vorhersagen über die Veränderung der Bauelementeeigenschaften, etwa über die Verkürzung der Schaltzeit, wenn die linearen Abmessungen des Bauelementes bei gleicher elektrischer Feldstärke, d. h. bei verkleinerter Betriebsspannung, reduziert werden. Es beruht auf bestimmten Transformationseigenschaften der mathematischen Gleichungen für die Beschreibung der Bauelemente. Nur solange diese Gleichungen gelten — oben haben wir Effekte genannt, die zum Teil über ihren Rahmen hinausgehen — ist das «scaling down» anwendbar. Es verliert seine Gültigkeit außerdem dann, wenn nichtskalierbare Randbedingungen wichtig werden. Zum Beispiel ist das Oberflächenpotential im Kanal eines MOS-Transistors von dessen Abmessun-

gen unabhängig. Bei Kanallängen unterhalb $1/4\text{ }\mu\text{m}$ wird der Einfluß statistischer Fluktuationen der Dotierungskonzentration im Kanal auf die Schwellspannung spürbar, d. h., die Schwellspannung beginnt von Transistor zu Transistor im Schaltkreis unkontrollierbar zu schwanken. Ein anderes Problem ist, daß die Zahl der Elektronen im Kanal so klein wird, daß die von radioaktiver Strahlung aus dem Schaltkreis oder aus der Umgebung erzeugten Elektronen unter Umständen ausreichen, den Transistor ein- oder auszuschalten. Die physikalischen Grenzen für die Verkleinerung der Halbleiterbauelemente und die Erzeugung der dafür notwendigen Strukturen durch eine Halbleitertechnologie sind heute gut untersucht. Vorerst, wenigstens solange die Strukturen $1/10\text{ }\mu\text{m}$ nicht unterschreiten, werden Grenzen spürbar, die nur relativen Charakter haben. Sie lassen sich durch konstruktive Veränderungen an den Bauelementen, durch die Wahl anderer Materialien (z. B. durch Metallsilizide als Gate-Elektrode) oder durch Betreiben der Schaltkreise bei tiefen Temperaturen, zunächst bei der des flüssigen Stickstoffs, weiter hinausschieben. Daraus kann geschlossen werden, daß die Mikroelektronik auf der Basis von Halbleiterbauelementen in der Lage sein wird, ausreichend schnelle Schaltkreise mit 10^7 bis 10^8 Bauelementen zu realisieren. Bei Integrationsgraden von 10^7 und 10^8 stellt sich aber auch die Frage nach dem Energieverbrauch neu. Oben hatten wir davon gesprochen, daß die Schaltkreise bereits energieeinsparend genug arbeiten. Diese Bemerkung bezog sich auf Schaltkreise mit einigen 10^5 Bau-

elementen. Aber auch bei 10^7 und 10^8 Bauelementen pro Schaltkreis besteht das Problem noch nicht darin, daß insgesamt zuviel Energie verbraucht wird, sondern darin, daß diese Energie in Wärme umgewandelt wird und eine zu starke Aufheizung des Schaltkreises bewirkt, wenn sie nicht schnell genug abgeleitet werden kann. Für die Wärmeableitung gibt es physikalische Grenzen. Abschätzungen zeigen, daß aus Schaltkreisgehäusen bei vertretbar hohen Temperaturen des Schaltkreises nicht mehr als etwa 10 Watt Wärme kontinuierlich abgeführt werden können. Daraus resultiert die Forderung, die Verlustleistung der Schaltkreise weiter zu reduzieren, z. B. durch den Übergang zu kleineren Spannungen. Grenzen sind hier u. a. durch die thermischen Spannungsschwankungen gesetzt, auf die wir schon hingewiesen haben. Die immer weitere Verkleinerung der Bauelemente bis heran an die prinzipiellen physikalischen Grenzen ist, wie wir wissen, kein Selbstzweck, sondern Mittel zum Zweck. Der eigentliche Zweck ist die Erhöhung des Integrationsgrades und der Arbeitsgeschwindigkeit der Schaltkreise. Zur Beherrschung höherer Integrationsgrade sind nicht nur die Probleme im Zyklus 1 zu lösen, auf die wir soeben eingegangen sind, sondern auch Probleme im Zyklus 0. Daß man Substratscheiben mit geringerer Defektdichte braucht, um größere Chips mit vertretbarer Ausbeute herstellen zu können, haben wir bereits erwähnt. Man sucht dieser Forderung durch Perfektionierung der einzelnen Schritte bei der Kristall- und Scheibenherstellung gerecht zu werden, ohne bisher je-

doch zu prinzipiell neuen Verfahren überzugehen. Zur Vorbehandlung der Scheiben für den Zyklus 1 werden in Zukunft in verstärktem Maße auch Getterungsverfahren eingesetzt werden, wodurch die unerwünschten Defekte aus der eigentlich aktiven dünnen Schicht auf der Vorderseite der Scheibe ins Innere oder auf die Rückseite verlagert werden, wo sie nicht stören. Man spricht von *Defekt-Engineering*. Die Vergrößerung der Durchmesser der Siliziumkristallstäbe und damit auch der Substratscheiben, auf 150 mm und darüber, die angestrebt wird, führt zwar nicht direkt zur Erhöhung der Integration in einem einzelnen Chip, vergrößert aber die Anzahl der Chips, die auf einer Scheibe untergebracht werden können und erhöht damit die Ausbeute im Zyklus 1. Indirekt ist die Vergrößerung des Scheibendurchmessers also doch eine Maßnahme zur Integrationserhöhung. Erhebliche Anforderungen stellt der Übergang zu höheren Integrationsgraden auch an den Funktionstest und an den Entwurf der Schaltkreise. Oft wird angenommen, daß der Aufwand für den Entwurf eines Schaltkreises exponentiell mit seinem Integrationsgrad steigt. Man versucht diesen enormen Anforderungen durch eine noch stärkere Inanspruchnahme der Computerunterstützung, aber auch durch die Entwicklung neuer Entwurfsprinzipien und Schaltkreiskonzepte gerecht zu werden. Ein Beispiel sind sich selbst testende Schaltkreise — die in jedem Schaltkreis ohnehin enthaltene Intelligenz wird für den Test mit genutzt. Wann Schaltkreise mit 10^7 und 10^8 Bauelementen hergestellt werden können, läßt sich heute noch nicht

mit Sicherheit vorhersagen. Gewiß sind die zu lösenden Schwierigkeiten größer und von grundsätzlicherer Natur als in der bisherigen Entwicklung der Mikroelektronik. Bisher konnte man davon ausgehen, daß sich der Integrationsgrad, der beherrscht wird, mit jedem Jahr verdoppelt. Nimmt man 1960 als das Jahr Null der Mikroelektronik mit einem Bauelement pro Schaltkreis, so lag danach 1970 der Integrationsgrad bei etwa $2^{10} \approx 10^3$, 1980 hätte er bei $2^{20} \approx 10^6$ liegen müssen, was aber noch nicht ganz der Fall war. Es ist unverkennbar, daß sich bereits Ende der siebziger, Anfang der achtziger Jahre eine Verlangsamung des Entwicklungstempos gegenüber den zurückliegenden knapp zwanzig Jahren abzeichnete. Ursachen dafür sind u. a. die qualitativ neuen Anforderungen an die Technologie, die Bauelementekonzeption, den Schaltkreisentwurf und den Funktionstest an der Schwelle zur Submikrometerelektronik, von denen wir oben gesprochen haben. Mitunter wird vermutet, daß die stärksten Beschränkungen nicht von der Technologie und den Bauelementen kommen werden, sondern vom Schaltkreisentwurf. Die Größe der Anforderungen, die beim Übergang zur Submikrometerelektronik zu bewältigen sind, lassen sich vielleicht wieder anhand der Autoindustrie verdeutlichen. Es ist etwa so ähnlich, als sollten von der Industrie Autos mit zehnfacher Höchstgeschwindigkeit und einem hundertfachen Fassungsvermögen entwickelt werden. Glücklicherweise stehen vor der Autoindustrie derartige Forderungen nicht. Sie wäre auch nicht in der Lage, sie zu erfüllen, es sei denn, sie würde

sich in die Luft begeben und das Flugzeug erfinden. Die Mikroelektronik steht vor der Aufgabe, neue integrierte Schaltkreise zu entwickeln, die sich zu den bisherigen wie das Flugzeug zum Auto verhalten und die trotzdem nicht wesentlich mehr kosten als die bisherigen. Sie scheint tatsächlich in der Lage zu sein, diese Aufgabe zu erfüllen, mit erheblichem Aufwand zwar und einschneidenden Veränderungen, aber ohne sich in ihrer heutigen Form als Halbleiterelektronik dabei selbst aufheben zu müssen.

Jedenfalls vorerst nicht. Der Spielraum der Mikroelektronik auf der Basis von Halbleitermaterialien ist noch sehr groß. Prinzipiell ist es z. B. nicht einzusehen, warum das Silizium das einzige Material sein sollte, das sich für die Herstellung integrierter Schaltkreise eignet und daß nicht Mittel und Wege gefunden werden sollten, um auch aus *anderen Halbleitermaterialien*, die in einzelnen Eigenschaften dem Silizium überlegen sind, integrierte Schaltkreise herzustellen. Vom Galliumarsenid war schon die Rede, das wegen seiner hohen Elektronenbeweglichkeit — $8800 \text{ cm}^2/\text{Vs}$ statt $1350 \text{ cm}^2/\text{Vs}$ beim Silizium — und aufgrund anderer Besonderheiten für die Mikroelektronik interessant ist. **Tatsächlich Galliumarsenid-Elektronik**

lich gibt es heute neben der Silizium-Mikroelektronik bereits eine gut entwickelte *Galliumarsenid-Mikroelektronik*. Wegen der ungünstigen Eigenschaften der Löcher des Galliumarsenid — z. B. beträgt deren Beweglichkeit nur $460 \text{ cm}^2/\text{Vs}$ — ist es eine rein *unipolare* Elektronik auf der Basis von Elektronen. Bipolare Transistoren kommen nicht zum Einsatz,

obwohl im Laboratorium auch daran gearbeitet wird. Das wichtigste Bauelement der Galliumarsenid-Elektronik ist der Feldeffekttransistor, der im Unterschied zum Silizium aber nicht als MOS-FET, sondern als MES-FET (von *Metal-Semiconductor-FET*) realisiert wird. Beim MES-FET (siehe Abb. 61a) liegt die metallische Gate-Elektrode direkt auf dem n-leitenden Galliumarsenid auf. Unter dem Gate bildet sich eine Verarmungsschicht aus, in die durch eine positive Spannung am Gate Elektronen hineingezogen werden können. Dadurch wird die Schicht leitend. Das Funktionsprinzip des MES-FET unterscheidet sich also etwas von dem des MOS-FET. Man spricht von einem *Sperrschicht-FET*. Damit sich der Beweglichkeitsvorteil des Galliumarsenid tatsächlich in kürzeren Verzögerungszeiten der MES-FETs niederschlagen kann, müssen Kanallängen bei oder unter $1 \mu\text{m}$ realisiert werden. MOS-FETs auf der Basis des Galliumarsenid sind u. a. deshalb problematisch, weil das Galliumarsenid kein ähnlich gutes Eigenoxid besitzt wie das Silizium. Daraus folgt auch, daß man für die Isolation von Leitbahnen ebenso wie für Passivierungs- und Maskierungszwecke in der Galliumarsenid-Elektronik auf andere Materialien zurückgreifen muß. Trotz dieser Schwierigkeiten wird die industrielle Herstellung von MES-FETs heute gut beherrscht. Als Einzelbauelemente kommen MES-FETs vor allem in Mikrowellenschaltungen zum Einsatz. Auch integrierte Schaltkreise aus MES-FETs sind auf dem Markt. Noch sind es solche mit kleinen und mittleren Integrationsgraden, darunter vor allem Analog-Schaltkreise für die Mikrowellen-

technik, aber auch Digital-Schaltkreise. Erste digitale LSI-Schaltkreise wurden 1980 im Laboratorium realisiert. Heute werden u. a. 1 KBit- und 4 KBit-RAM mit Zugriffszeiten bis zu 3 ns hergestellt. Für die hohen Arbeitsgeschwindigkeiten und geringen Verlustleistungen der Galliumarsenid-Schaltkreise (siehe Abb. 63) ist neben der hohen Elektronen-Beweglichkeit eine Eigenschaft des Galliumarsenid wichtig, die beim Silizium keine Entsprechung hat: Man kann Galliumarsenid auch als Isolator herstellen, z. B. durch Dotierung mit Chromatomen geeigneter Konzentration, vor allem aber durch Vermeidung jeglicher, auch unabsichtlicher Dotierung. Verwendet man (halb-)isolierendes Galliumarsenid als Substrat, werden die Maßnahmen zur gegenseitigen Isolation der Bauelemente überflüssig, die bei Silizium-Schaltkreisen unumgänglich sind und dort die Leistungsparameter negativ beeinflussen. Beim Silizium hilft man sich zum Teil mit fremden isolierenden Substraten, insbesondere aus Saphir (SOS-Technik).

Ein weiterer Vorteil des Galliumarsenid besteht darin, daß es sich mit anderen Verbindungshalbleitern, z. B. Aluminiumarsenid oder Galliumphosphid, gut mischen läßt. Durch geeignete Wahl des Mischungsverhältnisses kann man Halbleitermaterialien mit gewünschten elektrischen und optischen Eigenschaften herstellen — selbstverständlich innerhalb gewisser Grenzen. Solche «maßgeschneiderten» Halbleiter werden bereits seit längerem für die Herstellung optoelektronischer Bauelemente — Lumineszenzdioden, Laserdioden, Fotodetektoren u. a. — eingesetzt. Dadurch entsteht die Möglichkeit, auf

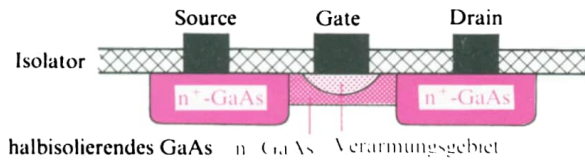
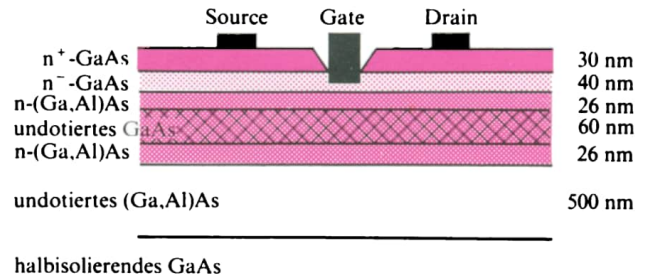


Abb. 61 Bauelemente der GaAs-Elektronik
MES-FET (a) und HEM-Transistor (b)



einem einzigen Chip aus Galliumarsenid sowohl elektronische Bauelemente, in erster Linie MES-FETs, als auch optoelektronische Bauelemente zu erzeugen.

Integrierte Mikrooptoelektronik

Erste Beispiele einer solchen integrierten Mikro-Optoelektronik wurden im Laboratorium bereits realisiert. Charakteristisch für die integrierten Chips der Mikro-Optoelektronik, aber auch bereits für die der Optoelektronik, ist die Verwendung sogenannter Halbleiter-Heterostrukturen. Darunter versteht man Schichtfolgen aus unterschiedlichen Halbleitermaterialien. Durch die Weiterentwicklung der epitaktischen Schicht-Abscheidungsverfahren, darunter vor allem der Molekularstrahlepitaxie (MBE) und der Metall-Organischen-CVD (MO CVD), ist es in den letzten Jahren gelungen, immer feinere und komplexere Heterostrukturen herzustellen — angefangen von Zweischichtstrukturen mit nahezu abruptem Übergang aus nur ein bis zwei Atomlagen über Mehrschichtstrukturen bis hin zu periodischen Vielschichtanordnungen aus zwei Materialien, den sogenannten *Supergittern*. Man spricht von *künstlichen Mikrostrukturen*. **Supergitter.**

Quantum Wells

An diesen Strukturen werden physikalische Effekte beobachtet, wie

man sie bei herkömmlichen Halbleitermaterialien und -strukturen nicht findet. Das in Abb. 61 b dargestellte Bauelement, der sog. HEMT (*High Electron Mobility Transistor*), beruht auf einem solchen Effekt — dem Einschluß von Elektronen in einer Materialschicht mit extrem hoher Beweglichkeit. In Abb. 61 b ist diese Schicht das undotierte Galliumarsenid. Die beiden daran angrenzenden Gallium-Aluminium-Arsenid-Schichten sind n-dotiert. Die von ihnen gelieferten Elektronen sammeln sich in der Galliumarsenid-Schicht wie in einer Mulde an. Wegen des Fehlens von Störatomen besitzen sie dort eine sehr hohe Beweglichkeit. Deshalb ist der HEMT ein sehr schnelles Bauelement — man erreicht Schaltzeiten von einigen ps ($1 \text{ ps} = 10^{-12} \text{ s}$). Die Elektronen beim HEMT können sich nur parallel zu den Wänden der Mulde, in die sie eingeschlossen sind, also in zwei Dimensionen, frei bewegen. Sie bilden ein *2-dimensionales Elektronengas*. Die Mulde selbst bezeichnet man als Quantum-Well. Eine völlig unerwartete Beobachtung am 2-dimensionalen Elektronengas ist der gebrochenzahlige *Quanten-Hall-Effekt*. Dabei verhalten sich die Elektronen so, als trügen sie nicht ein ganzes Elementarquantum «e» elektrischer Ladung, sondern nur einen Bruchteil davon.

Natürlich kann der Effekt nicht wirklich so interpretiert werden, denn die kleinste beobachtbare Ladung in der Natur ist und bleibt «e». Man muß die wahre Erklärung noch finden. Die Halbleitertheoretiker denken gegenwärtig darüber nach. Der *ganzzahlige* Quanten-Hall-Effekt bietet zwar keine prinzipiellen Verständnisschwierigkeiten, trotzdem ist auch er sensationell. Das Überraschende ist, daß sich mit seiner Hilfe Naturkonstanten, genauer das Verhältnis des Planck'schen Wirkungsquantums zur quadrierten Elektronenladung mit einer Genauigkeit von vorerst acht Stellen messen lassen — eine Genauigkeit, die kaum auf andere Weise zu erreichen ist. Entdeckt wurde der Quanten-Hall-Effekt 1980 von dem jungen Münchner Physiker K. von Klitzing. Insgesamt tut sich mit den Quantum Wells und den anderen künstlichen Mikrostrukturen ein breites Feld für die Grundlagenforschung in der Halbleiterphysik und die Anwendung ihrer Ergebnisse in neuen elektronischen und optoelektronischen Bauelementen auf. Durch die Einbeziehung künstlicher Mikrostrukturen in die Bauelementekonzepte und der Verfahren zu ihrer Erzeugung in die Herstellungstechnologie deutet sich gegenwärtig eine Weiterentwicklung der Mikroelektronik an, die von grundsätzlicher Natur ist. Sie bleibt

nicht auf das Galliumarsenid als Materialgrundlage beschränkt, sondern wirkt auch auf die Silizium-Mikroelektronik zurück und bezieht andere Verbindungshalbleiter, z. B. Indiumphosphid, mit ein. Das Indiumphosphid eignet sich in mancher Beziehung noch besser für die Mikro- und Optoelektronik als das Galliumarsenid und könnte sich in Zukunft sogar zu einem Konkurrenten für das Galliumarsenid entwickeln. Beide Materialien werden aber mit Sicherheit das Silizium nicht aus seiner dominierenden Stellung in der Mikroelektronik verdrängen. Eine Vorstellung von der tatsächlichen praktischen Bedeutung der Galliumarsenid-Elektronik im Vergleich zur Silizium-Elektronik vermitteln die folgenden Angaben: Lag 1975 die Zahl der pro Jahr verarbeiteten Galliumarsenid-Scheiben bei etwa 300 Tausend, so wird für 1985 ein Bedarf von etwa 1 Million abgeschätzt. Verglichen mit Silizium, wo bereits 1980 über 100 Millionen Scheiben pro Jahr verbraucht wurden, ist das immer noch eine bescheidene Zahl. Daran wird sich grundsätzlich auch in Zukunft nichts ändern. Ebenso wenig wie in der Metallindustrie der Stahl durch das Aluminium abgelöst wurde, wird in der Elektronik das Silizium durch das Galliumarsenid verdrängt werden. Die Galliumarsenid-Elektronik wird aber die Silizium-Elektronik wesentlich ergänzen und erweitern.

Funktionalelektronik

Auf der Basis von Galliumarsenid läßt sich auch ein Bauelement herstellen, das möglicherweise ein Vorbote für eine völlig andere Art von Elektronik, die sogenannte *Funktionalelektronik*, ist. Das Bauelement, von dem hier die Rede ist und das

als GUNN-Diode bezeichnet wird (nach dem Entdecker des Effektes), ermöglicht die direkte Erzeugung von elektromagnetischen Schwingungen im Mikrowellenbereich. Was üblicherweise durch Zusammenschaltung von Widerständen, Kondensatoren, Dioden und Transistoren erreicht wird, leistet ein einziges Bauelement. Die gewünschte Funktion, die Erzeugung von Schwingungen, wird durch die GUNN-Diode *unmittelbar* realisiert. Die Mikroelektronik zeichnet sich dadurch aus, daß die einzelnen Bauelemente zu einem Schaltkreis integriert werden. Das ist aber nur deshalb nötig, weil die Funktion, die der Schaltkreis erfüllen soll, vorher in die Einzelfunktionen von Bauelementen zerlegt wurde. Die Funktionalelektronik geht die Funktion direkt an. Die Integration wird noch eine Stufe höher getrieben als in der Mikroelektronik. Vorerst ist die Funktionalelektronik aber mehr ein Programm als eine klar erkennbare Entwicklungsrichtung der Elektronik. Ob sie jenes bleibt oder dieses werden wird, muß die Zukunft zeigen. Immerhin gibt es mit der sogenannten I^2L -Technik auch in der gegenwärtigen Siliziummikroelektronik dazu Ansätze.

3-dimensionale Integration

Von ähnlich grundsätzlichem Charakter für die Mikroelektronik wäre der Übergang von der Integration der Bauelemente in einer Ebene, die heute bis auf Ausnahmen durchweg verwendet wird, zur *Integration in mehreren übereinanderliegenden Ebenen*, also in den drei Dimensionen des Raumes. Die Zahl der Elemente pro Volumeneinheit ließe sich damit um ein Vielfaches steigern. Voraussetzung für die dreidimensionale In-

tegration ist u. a. die Beherrschung von Methoden, um auf der Passivierungsschicht, die eine bestimmte Schaltkreisebene abschließt, eine einkristalline Siliziumschicht für die Erzeugung der nächsten Schaltkreisebene abzuscheiden. An der Entwicklung solcher Methoden wird gegenwärtig gearbeitet. Besteht die Passivierungsschicht aus dem amorphen Siliziumdioxid, wie das heute fast ausschließlich der Fall ist, kann man z. B. die Graphoepitaxie benutzen. Bei dieser Methode werden in die amorphe Siliziumdioxidoberfläche feine Gräben eingeritzt. Wird auf der so behandelten Fläche Silizium abgeschieden, so wächst es ähnlich wie auf einem kristallinen Substrat als kristalline Schicht auf. Als aussichtsreich werden auch Verfahren angesehen, bei denen man vorhandene oder bewußt erzeugte Öffnungen in der Passivierungsschicht ausnutzt, um durch sie hindurch die Strukturinformation vom darunterliegenden Substrat nach oben auf die abzuscheidende Schicht übertragen zu können. Eine andere Methode besteht in der nachträglichen Ausheilung einer zunächst amorph abgeschiedenen Siliziumschicht ähnlich wie das im Anschluß an die Ionenimplantation beim gestörten Substrat geschieht. Auch kristalline Passivierungsschichten, z. B. aus Kalzium- und Strontiumdifluorid, auf denen das Silizium epitaktisch abgeschieden werden kann, werden erprobt.

Kryomikroelektronik

Die Palette von Halbleitermaterialien, die für die Mikroelektronik geeignet sind, kann dadurch erweitert werden, daß man zu *tiefen Temperaturen* übergeht. Materialien, wie

z. B. Indiumantimonid, die wegen ihrer relativ kleinen Energielücke zwischen Valenz- und Leitungsband bei Zimmertemperatur (273 K) eine zu große Ladungsträgerkonzentration besitzen oder deren Ladungsträgerbeweglichkeiten nicht hoch genug sind, nehmen bei der Temperatur des flüssigen Stickstoffs (77 K) oder des flüssigen Heliums (4 K) Eigenschaften an, die sie für die Mikroelektronik brauchbar machen. Deshalb werden heute Bauelemente und integrierte Schaltkreise, darunter auch solche aus Silizium oder Galliumarsenid, bei tiefen Temperaturen untersucht. Man spricht in diesem Zusammenhang von Kryomikroelektronik. Der Übergang zu tiefen Temperaturen hat u. a. auch den Vorteil, daß die Schwellspannungen der Transistoren kleiner werden, so daß mit kleineren Spannungspegeln und kleineren Verlustleistungen gearbeitet werden kann. Die Schnelligkeit der Bauelemente nimmt wegen der höheren Ladungsträgerbeweglichkeit bei tiefen Temperaturen ebenfalls zu.

Josephson-Elektronik

Eine radikalere Abkehr von der bisherigen Mikroelektronik, die ebenfalls tiefe Temperaturen erfordert, ist die sogenannte *Josephson-Elektronik*. Sie beruht auf einem von dem britischen Physiker B. Josephson 1962 entdeckten Effekt. Der Josephson-Effekt tritt an einer Festkörperanordnung auf, die aus zwei supraleitenden Metallschichten und einer dazwischenliegenden Isolatorschicht besteht (siehe Abb. 62). Wird die Isolatorschicht sehr dünn gemacht ($< 0,01 \mu\text{m}$), können die Elektronen aus den beiden Supraleitern durch sie hindurch «tunneln». Man spricht

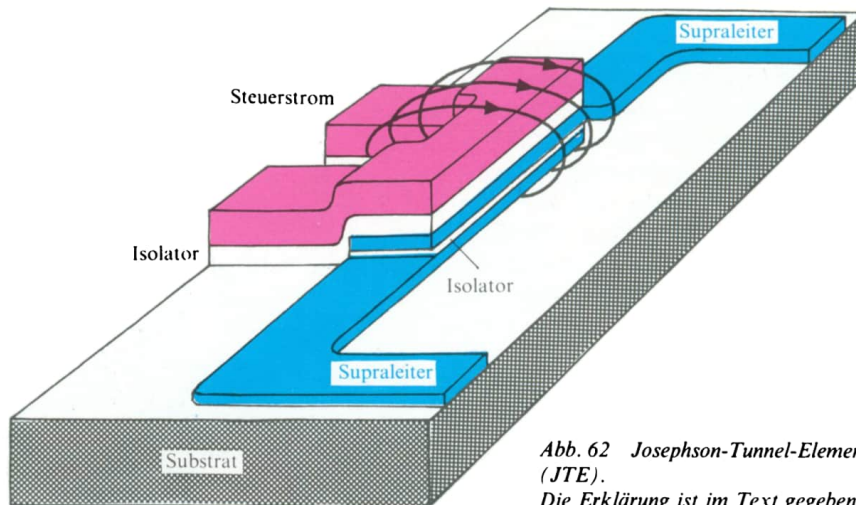


Abb. 62 Josephson-Tunnel-Element (JTE).

Die Erklärung ist im Text gegeben.

von einem Tunnelübergang. Da man sich die Gesamtheit der Elektronen in einem Supraleiter als makroskopische Welle mit einer bestimmten Phasenlage denken muß, kommt es beim Tunneln ähnlich wie bei der Überlagerung zweier kohärenter Lichtquellen zu Interferenzerscheinungen, in deren Ergebnis auch dann ein Strom durch den Übergang fließt, wenn überhaupt keine Spannung an ihn angelegt wird. Der Stromfluß durch einen Isolator ohne Spannung ist eine Merkwürdigkeit, die nur mit Hilfe der Quantentheorie der Supraleitung verstanden werden kann und deshalb der Anschauung zunächst schwer zugänglich ist. Nichtsdestoweniger ist der Effekt real und im Experiment gut nachweisbar. Bei Normalleitern besitzen die Elektronenwellen keine einheitliche Phase und die in beiden Richtungen durch den Übergang tunnelnden Elektronenwellen löschen sich gegenseitig durch Interferenz aus. Ohne Spannung fließt also durch den Übergang bei Normalleitung

kein Strom. Wenn ein Strom fließt, fällt auch eine (relativ große) Spannung am Übergang ab. Falls es gelingt, die beiden Leiter an einem Tunnelübergang zwischen dem supraleitenden und normalleitenden Zustand mit Hilfe einer kleinen Spannungsänderung hin- und herzuschalten und durch eine Stromquelle dafür zu sorgen, daß unabhängig vom Zustand ein konstanter Strom durch den Übergang fließt, so haben wir ein elektronisches Bauelement, das ähnlich wie ein Transistor Spannungsänderungen verstärkt. Das elektrische Umschalten zwischen Normal- und Supraleitung ist aber seit langem bekannt: Wird der Strom in einem Supraleiter über einen kritischen Wert hinaus erhöht bzw. wird der Supraleiter einem Magnetfeld oberhalb einer kritischen Stärke ausgesetzt, so bricht die Supraleitung in ihm zusammen und es stellt sich der normalleitende Zustand ein. Wird das Magnetfeld durch einen Strom erzeugt, der in einem zweiten Stromkreis fließt,

kann die große Spannung am Tunnelübergang durch die kleine Spannung in diesem zweiten Stromkreis gesteuert werden. Das ist es aber, was man für die digitale Informationsverarbeitung benötigt und worauf die Verwendung von Halbleiter-Transistoren in der Mikroelektronik beruht. Die beschriebene Anordnung bezeichnet man als Josephson-Tunnel-Element (JTE). Neben Tunnel-Elementen mit *einem* Übergang verwendet man auch Elemente mit zwei oder mehreren Übergängen, sogenannte SQIDs (Supraleitende Quanten-Interferenz Devices = Bauelemente). Ähnlich wie aus Halbleiterbauelementen lassen sich aus Josephson-Elementen Logik-Gatter, Speicher- und Rechenschaltkreise aufbauen. Die Schaltkreise können mit einer ähnlichen integrierten Plantechnik hergestellt werden, wie sie in der Halbleitermikroelektronik verwendet wird. Da ein Supraleiter dem Strom praktisch keinen Widerstand entgegengesetzt, verbrauchen die integrierten Schaltkreise aus Josephson-Elementen außerordentlich wenig elektrische Leistung. Allerdings darf bei einer Energiebilanz nicht außer acht gelassen werden, daß die Schaltkreise, bevor sie funktionieren, erst auf 4 K abgekühlt und auf dieser Temperatur gehalten werden müssen. Dafür ist eine Energie erforderlich, die den geringen Energieverbrauch beim Betreiben der Schaltkreise praktisch wieder kompensiert. Trotzdem ist die wesentlich geringere Verlustleistung der Josephson-Schaltkreise gegenüber den Halbleiterschaltkreisen von großer Bedeutung. Dadurch wird die Grenze der Integration, wie sie sich aus der Beschränkung der möglichen

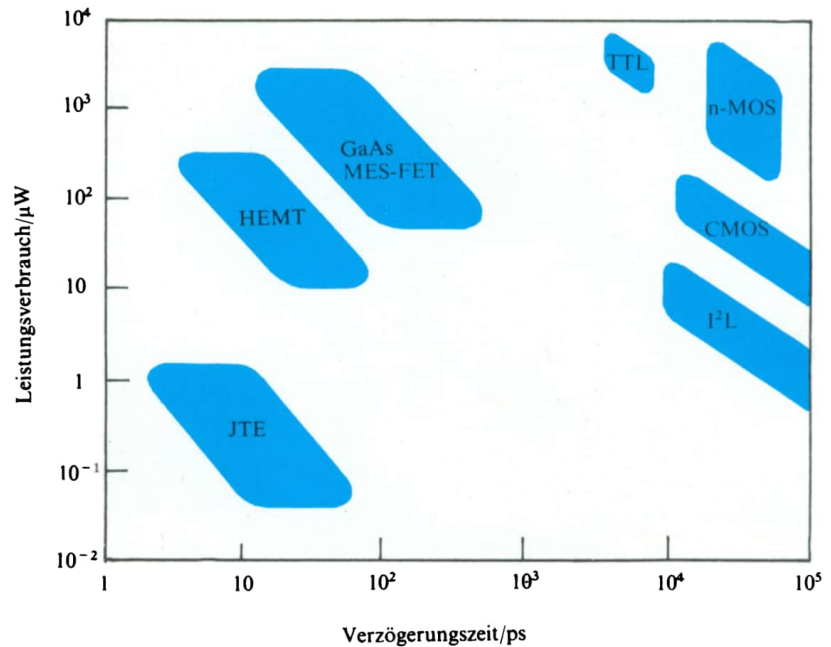


Abb. 63 Leistungsverbrauch und Verzögerungszeit von Logik-Gattern der GaAs- und Josephson-Elektronik. Zum Vergleich sind auch einige Logik-Gatter der Silizium-Elektronik dargestellt.

Wärmeabführung ergibt, weiter hinausgeschoben. Die vielen integrierten Schaltkreise, die für einen großen elektronischen Rechner nötig sind, können dadurch auf engstem Raum untergebracht werden. Daraus resultieren Vorteile für die Schnelligkeit der Rechner. Man muß sich dazu vergegenwärtigen, daß die Übertragung eines elektrischen Spannungsimpulses zwar mit Lichtgeschwindigkeit erfolgt, aber für eine Strecke von 30 cm immerhin noch eine Zeit von 1 ns erfordert. Mit Josephson-Schaltkreisen lassen sich Verzögerungszeiten realisieren, die bis zu 1/100 ns betragen können (siehe Abb. 63). Die Rechner müssen also sehr kompakt gebaut sein, damit die Schnelligkeit der Schaltkreise ausgenutzt werden kann. Die gerin-

gere Verlustleistung der Josephson-Schaltkreise macht eine solche kompakte Bauweise möglich. Von allen diskutierten Alternativen zur Halbleitermikroelektronik ist die Josephson-Elektronik ohne Zweifel die am weitesten entwickelte. Das ist auch deshalb bemerkenswert, weil das Wirkprinzip, der Josephson-Effekt, erst 1962 theoretisch vorhergesagt und ein Jahr später experimentell nachgewiesen wurde. Die theoretische Vorhersage beruhte übrigens auf der Übertragung von Erkenntnissen aus der Elementarteilchenphysik — dem Prinzip der Symmetriebrechung — auf die Festkörperphysik. Für die Entdeckung des Effektes erhielt Josephson 1973 den Nobelpreis für Physik zu einem Drittel. Heute gibt es, wie wir gese-

hen haben, bereits integrierte Schaltkreise auf der Grundlage dieses Effektes. Es ist immer wieder beeindruckend zu sehen, wie Ergebnisse der scheinbar abstraktesten Grundlagenforschung zu bedeutenden praktischen Auswirkungen führen können, und in welcher relativ kurzen Zeiträumen das geschehen kann. Allerdings ist die praktische Bedeutung, die die Josephson-Elektronik erlangt wird, gegenwärtig noch nicht klar abzusehen. Noch gibt es keine Josephson-Rechner, von sehr speziellen, nur im Labor realisierten, abgesehen. Die Herstellung der Josephson-Schaltkreise erweist sich als technologisch außerordentlich schwierig. Für die Verkleinerung der Josephson-Bauelemente zeichnen sich Grenzen ab, die sogar noch über denen für die Halbleiterbauelemente liegen.

Optische Bistabilität.

Handelt es sich bei den Josephson-Bauelementen immer noch um *elektronische* Bauelemente, d. h. um solche, die auf der Bewegung von Elektronen und deren Beeinflussung durch äußere elektrische und magnetische Einwirkungen beruhen, so führt eine andere moderne Richtung zur Entwicklung von Bauelementen für die Informationsverarbeitung gänzlich aus dem Gebiet der Elektronik heraus. Es handelt sich dabei um optische Bauelemente. Der Informationsträger ist das Licht. Durch Ausnutzung nichtlinearer optischer Effekte kann ein Halbleiter-Kristallplättchen mittels eines Laserstrahls zwischen den beiden Zuständen «lichtdurchlässig» und «undurchlässig» hin und hergeschaltet werden. Man bezeichnet diesen Effekt als *optische Bistabilität*. Die bei-

den Zustände «lichtdurchlässig» und «lichtundurchlässig» ermöglichen, ähnlich wie die Zustände «elektrisch leitend» und «elektrisch nicht leitend» in der Elektronik, prinzipiell die Entwicklung von Bauelementen für die digitale Informationsverarbeitung. Die auf dieses Ziel gerichteten Arbeiten befinden sich aber zur Zeit noch im Stadium der Grundlagenforschung. Es sollten sich Schaltzeiten im Pikosekundenbereich und Schaltleistungen im Milliwattbereich realisieren lassen, allerdings nicht beides gleichzeitig. Durch den Parallelbetrieb sehr vieler optischer Bauelemente könnte die Geschwindigkeit wesentlich gesteigert werden. Bereits Wirklichkeit ist die Ausnutzung des Lichtes als Informationsträger bei der Nachrichtenübertragung über Lichtleiter. Gegenwärtig wird daran gearbeitet, die für die Lichtleitertechnik verwendeten optoelektronischen Bauelemente, insbesondere die Halbleiterinjektionslaser, so weiterzuentwickeln, daß sie nicht nur für Übertragungszwecke, sondern auch für Aufgaben der optischen Signalverarbeitung eingesetzt werden können. 1983 berichteten japanische Wissenschaftler über den Bau eines Halbleiterlasers, dessen Lichtleistung durch einen zweiten, schwächeren Laser gesteuert werden kann. Es handelt sich dabei um ein ziemlich genaues Analogon zum Transistor, nur daß er statt mit elektrischen Ladungen mit Licht arbeitet, also ein *optischer* Transistor ist. Besondere Bedeutung für die **Integrierte Optik** optische Signalverarbeitung besitzen auch integrierte optische Strukturen aus Wellenleitern, Gittern, Modulatoren und anderen passiven optischen

Bauelementen. Strukturen der *integrierten Optik* wird man in Zukunft wahrscheinlich auch gemeinsam mit optoelektronischen und elektronischen Bauelementen auf einem einzigen Chip realisieren.

Molekularelektronik

Noch weiter in die Zukunft hinein reichen Überlegungen zur Nutzung chemischer und biologischer Prozesse für die Informationsverarbeitung. Denkt man daran, daß die Erbinformation bei den lebenden Organismen in den Molekülen der Desoxyribonukleinsäure gespeichert ist, so drängt sich die Frage auf, ob nicht auch eine technische Informationsverarbeitung möglich sein sollte, die mit Molekülen statt mit makroskopischen Objekten operiert, wie das heute durchweg der Fall ist. Ob in der Informationstechnik irgendwann auch einmal chemische oder biologische Hardware in Form von Molekülkomplexen eingesetzt werden wird, mag man bezweifeln oder nicht. Außer Frage steht, daß die Prinziplösungen für die Informationsverarbeitung, wie sie das Leben in Milliarden von Jahren entwickelt hat, auch für die Informationstechnik von Interesse sind.

Gesellschaftspolitische Faktoren

In Kapitel 2 haben wir beschrieben, was die Mikroelektronik zur Lösung eines der großen Probleme unserer Zeit, nämlich die Informationstechnik in den Griff zu bekommen, tun kann. Zweifellos wird das praktische Bedürfnis, Informationsprobleme aus allen Bereichen des Lebens effektiver als früher zu lösen und neue Informationsprobleme aufzugreifen und einer technischen Lösung zuzuführen, solange weiter bestehen, bis das letzte dieser Probleme zufrieden-

stellend gelöst ist. Dieser Zeitpunkt wird aber kaum jemals eintreten, weil sich die Gesellschaft und ihre materiellen Grundlagen selbst entwickeln werden und sich damit auch Informationsprobleme immer wieder neu stellen werden. Für *welche* Informationsprobleme anstelle der bisherigen technischen Lösungen die Mikroelektronik eingesetzt werden wird, welche Probleme überhaupt als technisch zu lösende angesehen werden, welche bewußt davon ausgeschlossen werden und welche Prioritäten gesetzt werden, all das hängt wesentlich von *gesellschaftspolitischen* Faktoren ab.

Eine entscheidende Rolle spielt die Mikroelektronik heute auch in der Militärtechnik. Experten schätzen ein, daß für militärische Anwendungen der Mikroelektronik etwa 20% der in der Welt insgesamt für wissenschaftliche Forschung eingesetzten Mittel aufgewendet werden. Wenn es gelingt, das Wettrüsten zu beenden, werden diese gewaltigen Mittel für die friedliche Nutzung der Mikroelektronik zur Verfügung stehen. Noch entscheidender ist aber, daß überhaupt nur in einer Welt, in der Frieden herrscht, auch die Perspektiven der Mikroelektronik, wie sie in diesem Abschnitt skizziert wurden, reale Gestalt annehmen können. Wenn von der künftigen Entwicklung der Mikroelektronik die Rede ist, können diese — hier nur angedeuteten — gesellschaftspolitischen Fragen nicht ausgeklammert werden. Gerade *weil* die Mikroelektronik so große Wirkungen in allen Bereichen der Gesellschaft erzielen kann, wirkt die Gesellschaft auch tiefgreifend auf sie zurück und bestimmt die Hauptlinien ihrer Ent-

wicklung entscheidend mit. In diesem Sinne ist die künftige Entwicklung der Mikroelektronik wesentlich mehr als eine Frage des technisch Machbaren.

Hinweise auf weiterführende Literatur

Völz, H.: Elektronik. Grundlagen, Prinzipien, Zusammenhänge. 2. Auflage. — Berlin: Akademie-Verlag 1979.

Möschwitzer, A.: Integration elektronischer Schaltungen und Mikrorechner. 3. Auflage. — Berlin: VEB Verlag Technik 1981.

Eckhardt, D.; Hadamowski, H.-F.; Jung-hans, B.; Schneider, H. G.: Mikroelektronik. Stand und Entwicklung. — Berlin: Akademie-Verlag 1982.

Schade, K.: Halbleitertechnologie, Bd. 1. — Berlin: VEB Verlag Technik 1981.

Schade, K. u. a.: Halbleitertechnologie, Bd. 2. — Berlin: VEB Verlag Technik 1983.

Möschwitzer, A.: Grundkurs Mikroelektronik. Standardschaltkreise, Kundenschaltkreise. — Berlin: VEB Verlag Technik 1984.

Kühn, E.; Schmied, H.: Handbuch integrierter Schaltkreise. 2. Auflage. — Berlin: VEB Verlag Technik 1980.

Probleme der Festkörperelektronik, Bd. 9—12. — Berlin: VEB Verlag Technik 1979—82.

Rost, A.: Grundlagen der Elektronik. — Berlin: Akademie-Verlag 1983.

Weißmantel, Ch.; Hamann, C. u. a.: Grundlagen der Festkörperphysik. 2. Auflage. — Berlin: VEB Deutscher Verlag der Wissenschaften 1981.

Bonč-Bruevič, V. L.; Kalašnikov, S. G.: Halbleiterphysik. — Berlin: VEB Deutscher Verlag der Wissenschaften 1982.

Völz, H.: Information I und II, 1. Auflage. — Berlin: Akademie-Verlag 1982.

Schwarz, W.; Meyer, G.; Eckhardt, D.: Mikrorechner. — Berlin: VEB Verlag Technik 1980.

Kieser, H.; Meder, U.: Mikroprozessor-technik. — Berlin: VEB Verlag Technik 1982.

Franke, K.: Einführung in die Mikrorechner-technik. — Berlin: VEB Verlag Technik 1984.

Hesse, S.: Kleines Lexikon der Industrie-robotertechnik. — Berlin: VEB Verlag Technik 1984.

Dubrau, J.; Netzschwitz, W.: Mikroelektronik. Wie verändert sie unser Leben? — Berlin: Dietz-Verlag 1983.

Nachweis der Abbildungen

Ein Teil der fotografischen Bildvorlagen wurde durch den VEB Kombinat Mikroelektronik und die DEWAG Leipzig zur Verfügung gestellt. Für einige Abbildungen wurden Vorlagen aus den Zeitschriften Radio Fernsehen Elektronik, VEB Verlag Technik Berlin; Wissenschaft und Fortschritt, Akademie-Verlag Berlin; URANIA, URANIA-Verlag Leipzig/Jena/Berlin sowie aus den Buchpublikationen H. Völz, Elektronik, Akademie-Verlag Berlin und Proceedings of the 16th Internat. Conf. on the Physics of Semiconductors, North-Holland Publ. Co. Amsterdam benutzt.

Inhaltsverzeichnis

Vorwort 5

1. Elektronik 9

- Elektronen
- Elektrische Energieumwandlung
- Leistungselektronik und Informationselektronik

2. Mikroelektronik, Information und Gesellschaft 12

- Informationsprozesse in der Gesellschaft. Gewinnung, Speicherung, Verarbeitung und Übertragung von Information
- Informationsbegriff. Information und Biologie
- Informationsverarbeitung mit Hilfe der Mikroelektronik
- Besser als frühere technische Lösungen der Informationsverarbeitung
- Technische Lösungen für früher unlösbare Probleme. Automatisierung
- Auswirkungen auf andere Informationsprozesse. Neue Kommunikationstechnologien
- Einheitliches Informationssystem
- Mikroelektronik und wissenschaftlich-technische Revolution

3. Halbleiterbauelemente 27

- Halbleiter
- Energiebänder
- Elektronen im Leitungsband und Löcher im Valenzband als frei bewegliche Ladungsträger
- Dotierung. Verunreinigungen und Strukturdefekte
- Flache und tiefe Niveaus
- Erzeugung von Überschußladungsträgern. Rekombination

- pn-Übergang. Halbleiterdiode. Injektion
- Bipolarer Transistor
- Feldeffekttransistor. MOS-FET
- Widerstände. Kondensatoren
- Optische Anzeigebaulemente. Halbleitersensoren

4. Informationsverarbeitung mit Hilfe elektronischer Schaltungen 42

- Analoge und digitale Informationsverarbeitung
- Binärzahlen
- Elementare logische Operationen
- Schaltungen für die elementaren logischen Operationen. Logik-Gatter. Schaltungsfamilien
- Speicherschaltungen. Flipflop-Schaltung. Kondensatorschaltung
- Adressierung von Speicherzellen
- Zählschaltungen
- Codier- und Decodierschaltungen
- Multiplexer-Schaltung
- Addierschaltungen
- Analogschaltungen. Analog-Digital-Wandler

5. Herstellungstechnologie der Mikroelektronik 61

- Integrierte Schaltkreise. Planartechnik
- Zyklus 1: Oxydation. Strukturierung durch Fotolithografie. Ätzen. Dotierung. Epitaxie. Metallisierung. Der Gesamtprozeß am Beispiel. Integrationsgrade. Realisierungstechniken. Reinheit und Präzision
- Zyklus 0: Schaltkreisentwurf. Foto-schablonen. Substratscheiben
- Zyklus 2: Vereinzeln der Plättchen. Bonden. Verkappen
- Herstellungskosten

6. Integrierte Schaltkreise und ihre Anwendung 78

- Ein passender Schaltkreis für jedes Problem?
- Problemkomplexität und Integrationsgrad. Problemvielfalt und Programmierung. Problemhäufigkeit und Stückzahl
- Kundenwunsch- und Standardschaltkreise
- Uhrenschaltkreis. Taschenrechner-schaltkreis. Kameraschaltkreis. Analoge Schaltkreise
- Logik-Gatter. MSI-Standardschaltkreise
- Speicherschaltkreise: ROM, RAM, EPROM
- Mikroprozessoren
- Mikrorechner

7. Perspektiven der Mikroelektronik 99

- Intelligentere Informationssysteme. Höchstmögliche maschinelle Intelligenz zu einem niedrigen Preis. Höhere Integrationsgrade und höhere Arbeitsgeschwindigkeit
- Submikrometer-Technologie. Submikrometer-Baulemente
- Galliumarsenid-Elektronik. Integrierte Mikrooptoelektronik
- Mikrostrukturen. Supergitter. Quantum Wells
- Funktionalelektronik
- 3-dimensionale Integration
- Kryomikroelektronik
- Josephson-Elektronik
- Optische Bistabilität. Integrierte Optik
- Molekularelektronik
- Gesellschaftspolitische Faktoren

Hinweise auf weiterführende Literatur 115

