

Universität Stuttgart
Fakultät 5

Prof. Dr.-Ing. U. G. Baitinger

Elektrotechnische Grundlagen

Diplomstudiengang Informatik
1. Semester

1 Grundgrößen und Grundregeln

1.1 Elektrische Ladung

1.2 Elektrische Spannung

Elektrische Feldstärke
Potential und Spannung
Ideale und reale Spannungsquellen
Die Kirchhoffsche Maschenregel

1.3 Elektrischer Strom

Magnetische Feldstärke
Stromstärke und Stromdichte
Ideale und reale Stromquellen
Die Kirchhoffsche Knotenregel

1.4 Elektrischer Widerstand

Das Ohmsche Gesetz

1.5 Energie im Stromkreis

Elektrische Arbeit und Leistung

2 Passive Bauelemente

2.1 Darstellungen mit Zweipolen

Elektrische Zweipole
Elektrische Stromkreise

2.2 Resistive Zweipole

Elektrischer Widerstand
Technische Widerstände
Schaltungen mit idealen Widerständen

2.3 Induktive Zweipole

Elektrische Induktivität
Technische Spulen
Schaltungen mit idealen Induktivitäten

2.4 Kapazitive Zweipole

Elektrische Kapazität
Technische Kondensatoren
Schaltungen mit idealen Kapazitäten

2.5 Nichtlineare Zweipole

Silizium-Dioden

3 Lineare passive Schaltungen

3.1 Das Überlagerungsprinzip

3.2 Ausgleichsvorgänge im Gleichstromkreis

Einschalt- und Abklingvorgänge von
Spulenstrom und Kondensatorspannung

3.3 Reihenschaltungen im Wechselstrombetrieb

verlustbehaftete Spule
verlustbehafteter Kondensator
Grenz- und Resonanzfrequenz
Reihenschwingkreis

3.4 Parallelschaltungen im Wechselstrombetrieb

verlustbehafteter Kondensator
verlustbehaftete Spule
Parallelschwingkreis

3.5 RC-Glieder im Impulsbetrieb

Differenzierglied
Integrierglied

4 Stromleitung in Halbleitern

4.1 Grundlagen des Halbleitereffekts

Kristallgitter und Gitterstörungen
Energiebänder und Energieterme
Eigenleitung
n-Leitung mit Donatoren
p-Leitung mit Akzeptoren
Kompensation

4.2 Dichte der Ladungsträger

Eigenleitungsdichte
Störleitungsdichte

5 Aktive Bauelemente

5.1 Der PN-Übergang

Das thermische Gleichgewicht

Durchlaß- und Sperrbereich

5.2 Bipolare Transistoren

Aufbau und Wirkungsweise

Kennlinien und Betriebsbereiche

5.3 Der MOS-Kondensator

Anreicherungstyp

Verarmungstyp

5.4 MOS-Transistoren

Aufbau und Wirkungsweise

Kennlinien und Betriebsbereiche

7 Digitale Grundsaltungen

7.1 Inverterschaltungen

Kontaktdarstellung von Transistoren

Lastwiderstand und Lasttransistor

Transistoren im Schalterbetrieb

7.2 Elektronische Verstärkung

Verstärkerbetrieb mit kleinen Signalen

Digitalbetrieb mit großen Signalen

7.3 Transistoren im Schalterbetrieb

Bipolare Transistoren als Schalter

MOS-Transistoren als Schalter

6 Mikroelektronik

6.1 Ersatzschaltbilder der Mikroelektronik

Aktive Bauelemente

Passive Bauelemente

6.2 Aufbau integrierter Bauelemente

Bipolare Transistoren

MOS-Transistoren

Widerstände, Kondensatoren, Dioden

8 Verknüpfungsschaltungen

8.1 Realisierung von Schaltfunktionen

8.2 Logik mit bipolaren Transistoren

Direkt gekoppelte Transistor-Logik (DCTL)

Dioden-Transistor-Logik (DTL)

Transistor-Transistor-Logik (TTL)

Emittergekoppelte Logik (ECL)

8.3 Logik mit MOS-Transistoren

Statische NMOS-Logik

Statische PMOS-Logik

Statische CMOS-Logik

Dynamische Domino-Logik

9 Speicherschaltungen

9.1 Speicherzellen mit MOS-Transistoren

Statische Speicherzelle

Dynamische Speicherzellen

9.2 Speichermatrizen (RAM)

Blockschaltbild

Chip-Layout

10 Digitale Schaltungsstrukturen

10.1 Technologie und Schaltungstechnik

Technologische Randbedingungen

Schaltkreistechnologien

Allgemeines Strukturschema

10.2 Programmierbare Schaltungen

Festwertspeicher (ROM)

Programmierbare UND-Matrixlogik (PAL)

Programmierbare Logische Matrix (PLA)

10.3 Integrierte Schaltungen

Integrierte Schaltungsstrukturen

Standard Logik-ICs

Voll-kundenspezifische Logik-ICs

Semi-kundenspezifische Logik-ICs

Literatur

Elektrotechnik und Elektronik für Informatiker/innen

Band 1 Grundbegriffe der Elektrotechnik

Band 2 Grundgebiete der Elektronik

von Prof. Dr.-Ing. Reinhold Paul

Technische Universität Hamburg-Harburg



B. G. Teubner Verlag Stuttgart 1995

Für elektrische Vorgänge besitzt der Mensch von Natur aus kein Sinnesorgan; er kann sie nur indirekt durch optische, akustische oder thermische Erscheinungen wahrnehmen. Die Messung elektrischer Größen erfolgt daher stets durch ihre Umwandlung in eine wahrnehmbare physikalische Größe, z.B. einen Zeigerausschlag. Das heißt, es wird nicht die Elektrizität an sich, sondern es werden ihre Wirkungen festgestellt, wie Anziehungs- oder Abstoßungskräfte, Leuchterscheinungen oder die Erwärmung materieller Körper. Die qualitative Erklärung der Auswirkungen der Elektrizität erfolgt durch physikalische Modelle, deren Gültigkeit prinzipiell Grenzen gesetzt sind, d.h. sie gelten nur zur Erklärung des gerade betrachteten Sachverhalts. Ihre quantitative Erfassung gelingt mit mathematischen Methoden; das Handwerkszeug des Elektrotechnikers sind daher spezielle mathematische Verfahren. - Hier erfolgt die Betrachtung elektrischer Vorgänge, ihre Erklärung mit Modellen und ihre Erfassung mittels mathematischer Gleichungen im Hinblick auf ihre technische Nutzung in Bauelementen und Systemen der Informationstechnik.

1.1 Elektrische Ladung

Die elektrische Ladung ist eine physikalische Grundgröße; man bezeichnet sie mit dem Zeichen Q . Ein Elektron besitzt die kleinste elektrische Ladung, die „Elementarladung“; sie beträgt

$$e = -(1,602 \cdot 10^{-19})\text{C}$$

$$Q = N \cdot e \quad [Q] = \text{C} \quad (1.1)$$

Die Dimension $[Q]$ der Ladung heißt C(oulomb), benannt nach dem französischen Physiker und Ingenieur *Charles Augustin Coulomb* (1736-1806), der als erster festgestellt hat, daß es zwei entgegengesetzte Arten elektrischer Ladung gibt, die er als „positiv“ bzw. „negativ“ bezeichnete, und daß sich ungleichartige Ladungen anziehen, während sich gleichartige abstoßen.

1.2 Elektrische Spannung

1.2.1 Die elektrische Feldstärke

Man beobachtet, daß elektrische Kräfte über große Entfernungen und auch ohne Übertragungsmedium wirken. Man bedient sich daher der Modellvorstellung des „elektrischen Feldes“. Es besteht in der Umgebung einer ruhenden Ladung bzw. zwischen zwei ruhenden Ladungen (oder elektrisch geladenen Körpern). Beim elektrischen Feld liegt eine doppelte Wechselwirkung vor:

- die Kraftwirkung, die ein vorhandenes elektrisches Feld auf eine ruhende Ladung ausübt, und
- eine ruhende Ladung als Ursache eines elektrischen Feldes.

Das elektrische Feld einer ruhenden Ladung vermag auf eine andere ruhende Ladung Q eine Kraft F_e auszuüben, die bekanntlich eine gerichtete Größe („Vektor“) ist; die „elektrische Feldstärke“ E ist daher ebenfalls ein Vektor:

$$\vec{F}_e = Q \cdot \vec{E} \quad (1.2)$$

Der Vektor der elektrischen Feldstärke E ist gleich gerichtet wie der Vektor der wirkenden Kraft F_e , da die Ladung Q ein Skalar ist. Elektrische Feldlinien beginnen und enden auf Ladungen; sie zeigen per definitionem von positiven Ladungen weg und zu negativen Ladungen hin. Damit lassen sich auch Anziehung und Abstoßung wie in Bild 1.1 veranschaulichen.

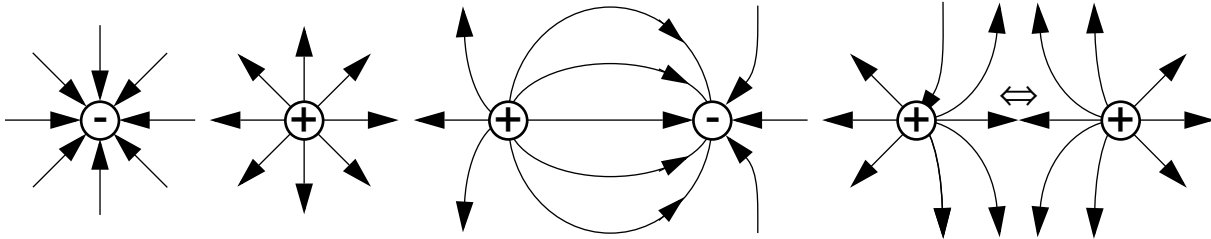


Bild 1.1: Elektrische Feldlinien ruhender Ladungen; - negative, + positive Ladungen; links: einzelne Ladungen, mitte: Anziehung, rechts: Abstoßung.

Umschließt in einem elektrischen Feld eine geschlossene Hüllfläche A eine elektrische Ladung Q (bzw. einen oder mehrere elektrisch geladene Körper), so gilt:

$$\oint \vec{D} d\vec{A} = Q \quad \quad \quad \oint \vec{D} d\vec{A} = 0 \text{ für } Q = 0 \quad (1.3)$$

(Das Umlaufintegral verschwindet zwangsläufig, wenn keine Ladungen umschlossen werden.) Mathematisch wird ein solches Integral als „Fluß“ bezeichnet, so daß man bei der Größe D von einer „elektrischen Flußdichte“ sprechen kann. Sie ist die Ursache des elektrischen Feldes; dessen Wirkung war nach Gl.(1.2) die elektrische Feldstärke E . Beide Größen sind zueinander proportional:

$$\vec{D} = \epsilon \cdot \vec{E} = \epsilon_r \epsilon_0 \cdot \vec{E} \quad \quad \quad \epsilon_0 = 0,885 \cdot 10^{-11} \text{ As/Vm} \quad (1.4)$$

Proportionalitätsfaktor ist die „Dielektrizitätskonstante“ ϵ , wobei ϵ_0 für das Vakuum gilt, während die relative Dielektrizitätszahl ϵ_r vom Material abhängt, in dem das elektrische Feld verläuft. Bei gleicher Flußdichte D ergeben sich je nach Material stark unterschiedliche Feldstärken E und damit Kraftwirkungen des elektrischen Feldes:

- Luft: $\epsilon_r = 1$; Polypropylen: $\epsilon_r = 2,2$; Quarzglas (SiO_2): $\epsilon_r = 3,9$; Silizium (Si): $\epsilon_r = 12$.

1.2.2 Das elektrische Potential

Zur Trennung von Ladungen ist Energie erforderlich; der Energiezustand einer Ladung ist ihre potentielle Energie, kurz ihr „Potential“. Falls es sich um eine ruhende Ladung handelt, so ist jeder Punkt ihres elektrischen Feldes gleichberechtigt durch den Vektor der elektrischen Feldstärke E und den Skalar des elektrischen Potentials φ gekennzeichnet, wobei gilt:

$$\varphi = \int \vec{E} d\vec{s} + \text{const} \quad \quad \quad \oint \vec{E} d\vec{s} = 0 \quad (1.5)$$

Integriert man über einen geschlossenen Umlauf, so verschwindet die potentielle Energie. Zwischen zwei räumlich getrennten Ladungen in den Punkten a und b im Abstand s besteht nach Gl.(1.5) ein Potentialunterschied, der als (elektrische) „Spannung“ und mit dem Zeichen U bezeichnet wird:

$$U_{ab} = \varphi_a - \varphi_b = \int_a^b \vec{E} d\vec{s} = \frac{1}{Q} \int_a^b \vec{F}_e d\vec{s} = \frac{W}{Q}$$

Dabei ist Q die betrachtete Ladung, W die elektrische Energie, die für die Ladungstrennung aufgewendet werden muß. Bei einer einzelnen Ladung kann die unendlich weit entfernte Potentialfläche als Bezugspotential φ_0 definiert werden, z.B. das der Erde („Masse“). Man erhält so die Bestimmungsgleichung für die elektrische Spannung, die ebenfalls eine physikalische Grundgröße ist:

$$U = \varphi - \varphi_0 = \frac{W}{Q} \quad [U] = V \quad (1.6)$$

Die Dimension $[U]$ der Spannung heißt V(olt), benannt nach dem italienischen Physiker *Allessandro Volta* (1745-1827), Erfinder des Elektrometers und der Voltaschen Säule zur Spannungserzeugung.

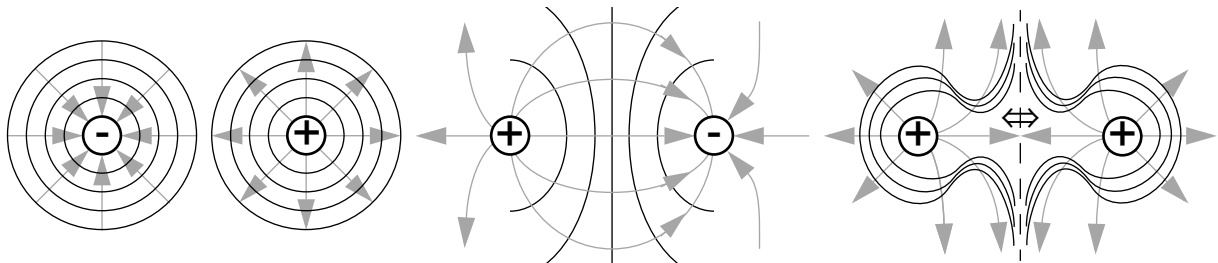


Bild 1.2: Äquipotentiallinien ruhender Ladungen

In der Modellvorstellung des elektrischen Feldes verlaufen die Flächen gleichen Potentials um punktförmige Ladungen konzentrisch, sind also Kugelflächen, deren Flächennormalen die Vektoren der elektrischen Feldstärke sind, wie Bild 1.2 zeigt.

1.2.3 Spannungsquellen

Bild 1.3 zeigt links zwei verschiedene Symbole, die zur Darstellung idealer Gleichspannungsquellen verwendet werden; U_0^e ist die „eingeprägte“ Spannung, die von der Quelle geliefert wird. Reale Spannungsquellen sind mit Energieverlusten behaftet, die ersatzweise durch einen Innenwiderstand R_i dargestellt werden können; sie geben nicht den vollen Wert der eingeprägten Spannung an den Verbraucher ab, hier einen Widerstand R .

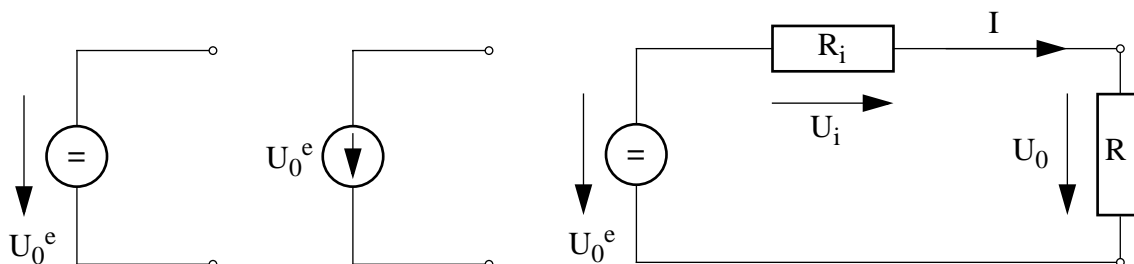


Bild 1.3: Ideale und reale Gleichspannungsquellen

1.2.4 Die Kirchhoffsche Maschenregel

Schaltet man mehrere elektrische Bauelemente in Reihe, wie Bild 1.4 (links) am Beispiel von Widerständen zeigt, so fließt durch alle derselbe Strom. Die Teilspannungen, die an den Widerständen abfallen, addieren sich zum Wert der angelegten Spannungsquelle U_0 . Allgemein ausgedrückt ist in einem beliebigen Netzwerk aus zweipoligen Erzeugern und Verbrauchern bei einem geschlossenen Umlauf, einer sogenannten *Masche*, die *Summe der Spannungen gleich Null*:

$$U_0 = U_1 + U_2 + U_3 \qquad -U_0 + U_i + U_L + U + U_A = 0 \quad (1.7)$$

Diese Grundregel, die nach *Gustav Robert Kirchhoff* (1824-1887) benannt wurde, ist unmittelbar einsichtig, da außer den betrachteten Spannungen - seien es solche aus Spannungsquellen oder an Verbrauchern abfallende - keine weiteren im betrachteten Umlauf vorliegen. Die „Kirchhoffsche Maschenregel“ ist ein Axiom der Elektrotechnik, d.h. sie ist einfach, plausibel und widerspruchsfrei und sie bedarf keines weiteren Beweises.

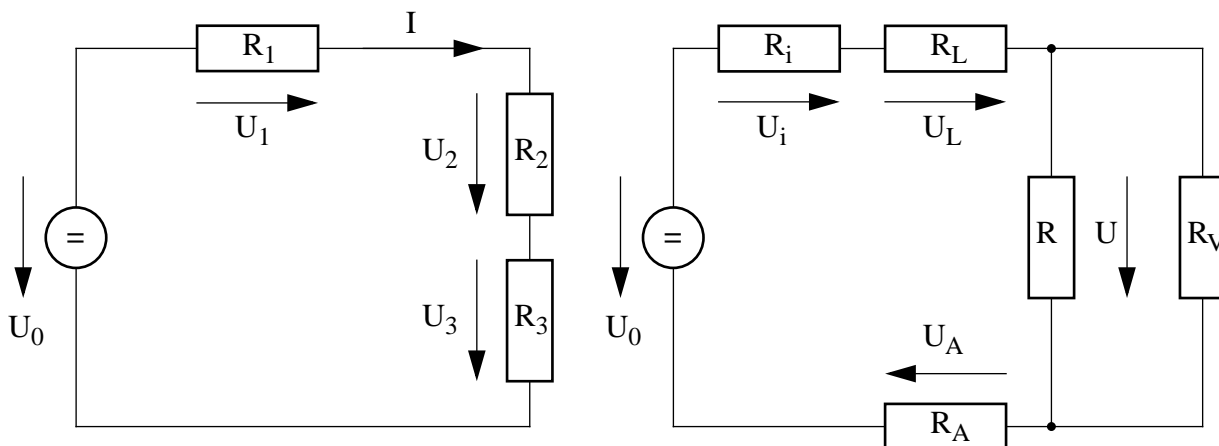


Bild 1.4: Zur Kirchhoffschen Maschenregel

Man kann sich darauf einigen, Spannungspfeile im rechtsdrehenden Umlaufsinn als positiv, die entgegengesetzten als negativ anzusetzen. Positive Spannungspfeile an Verbrauchern und negative Spannungspfeile an Erzeugern (= Spannungsquellen) entsprechen dann Spannungswerten > 0 .

1.3 Elektrischer Strom

1.3.1 Die elektrische Stromstärke

Bewegte elektrische Ladungen werden als elektrischer Stromfluß oder kurz als (elektrischer) „Strom“ bezeichnet. Als Ladungsträger kommen einzelne Elektronen oder geladene Atome in Betracht. Prinzipiell ist die Bewegung von Ladungsträgern sowohl im Vakuum als auch im stoffgefüllten Raum möglich, wobei alle drei Aggregatzustände der Materie in Frage kommen:

- Stromleitung durch quasi-freie Elektronen oder Defektelektronen in Festkörpern (Kristallen);
- Bewegungen von Ionen durch Flüssigkeiten (z.B. in Elektrolyten);
- Gasentladung durch Ionen in einem Plasma (z.B. in Leuchtstoffröhren).

Technisch interessant und in der elektronischen Schaltungstechnik nutzbar ist vor allem der Stromfluß in metallischen Leitern und in Halbleitern, wie z.B. Silizium. Metallatome geben bereits ohne Energiezufuhr die Elektronen ihrer äußersten Elektronenschale ab, die sich dann wie Gasmoleküle quasi-frei durch das Kristallgitter bewegen können, während die Halbleiter hierzu der Energiezufuhr bedürfen, wozu bei Silizium bereits die Zimmertemperatur ausreicht. Isolatoren setzen dagegen selbst bei hoher Energiezufuhr keine nutzbare Anzahl von Elektronen frei. Unter dem Einfluß einer von außen an einen Leiter angelegten elektrischen Spannung beobachtet man eine Vorzugsrichtung der quasi-freien Elektronen, die sich gemittelt vom Anschluß mit Elektronenüberschuß („Minuspole“) zum Anschluß mit Elektronenmangel („Pluspol“) bewegen. Die Elektronen bewegen sich also vorzugsweise entlang eines *Potentialgefälles*.

Der elektrische Strom, d.h. die gerichtete Bewegung elektrischer Ladungen, ist wie die Ladung selbst eine physikalische Grundgröße, für die man das Zeichen I verwendet. Die Bestimmungsgleichung lautet:

$$I = \frac{Q}{t} = \frac{N \cdot e}{t} \quad [I] = A \quad (1.8)$$

wobei Q die Ladung, e die Elementarladung, N die Anzahl der Elementarladungen und t die Zeit ist. Die Dimension $[I]$ des Stromes heißt A(mpere), benannt nach dem französischen Physiker und Mathematiker *André-Marie Ampère* (1775-1836), dem Begründer der elektrodynamischen Theorie.

1.3.2 Die elektrische Stromdichte

Bei gleichmäßiger („homogener“) Verteilung des Stromes I über den Querschnitt A eines metallischen Leiters läßt sich die Stromdichte, die mit dem Zeichen J bezeichnet wird, einfach berechnen:

$$J = \frac{I}{A} \quad [J] = [I]/[A] = A/mm^2 \quad (1.9)$$

Was die Dimension $[J]$ der Stromdichte betrifft, so verwechsle man bitte nicht die Dimension $[I] = A(\text{mpere})$ mit der Querschnittsfläche A , deren Dimension $[A] = mm^2$ ist.

1.3.3 Stromquellen

Bild 1.5 zeigt links zwei verschiedene Symbole, die zur Darstellung idealer Gleichstromquellen verwendet werden; I_0^e ist der „eingeprägte“ Strom, der von der Quelle geliefert wird. Reale Stromquellen sind mit Energieverlusten behaftet, die ersatzweise durch einem Innenleitwert $G_i = 1/R_i$ dargestellt werden können; sie geben nicht den vollen Wert des eingepprägten Stroms an den Verbraucher ab, hier einen Widerstand R .

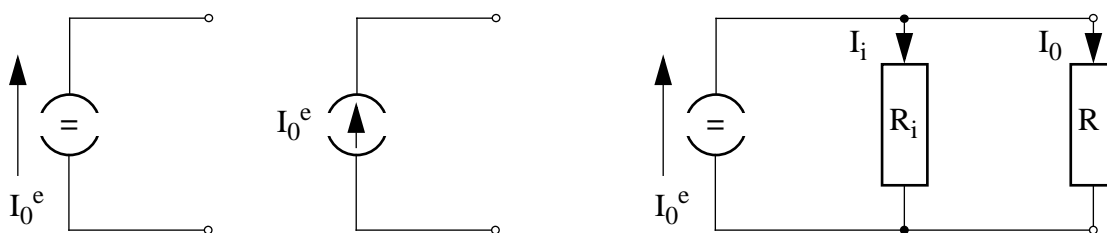


Bild 1.5: Ideale und reale Gleichstromquellen

1.3.4 Die Kirchhoffsche Knotenregel

Schaltet man mehrere elektrische Bauelemente parallel, wie Bild 1.6 (links) am Beispiel von Widerständen zeigt, so liegen alle an derselben Spannung. Die Teilströme, die durch die Widerstände fließen, addieren sich zum Wert der angelegten Stromquelle I_0 bzw. zum Gesamtstrom, den eine angelegte Spannungsquelle U_0 liefert. Allgemein ausgedrückt ist für jeden *Knoten* in einem beliebigen Netzwerk aus zweipoligen Erzeugern und Verbrauchern die *Summe der Ströme gleich Null*:

$$I_0 = I_1 + I_2 + I_3 \qquad -I_A + I + I_V = 0 \qquad (1.10)$$

Auch diese Grundregel ist unmittelbar einsichtig, da an dem betreffenden Knoten außer den betrachteten Strömen keine weiteren zu- oder abfließen. Die „Kirchhoffsche Knotenregel“ ist daher ebenfalls ein Axiom der Elektrotechnik, das keines weiteren Beweises bedarf; auch sie ist einfach, plausibel und widerspruchsfrei. Man kann sich darauf einigen, zuführende Strompfeile als positiv, wegführende als negativ anzusetzen. Die beiden Kirchhoffschen Regeln, zusammen betrachtet, zeigen die *Dualität* von Strom und Spannung.

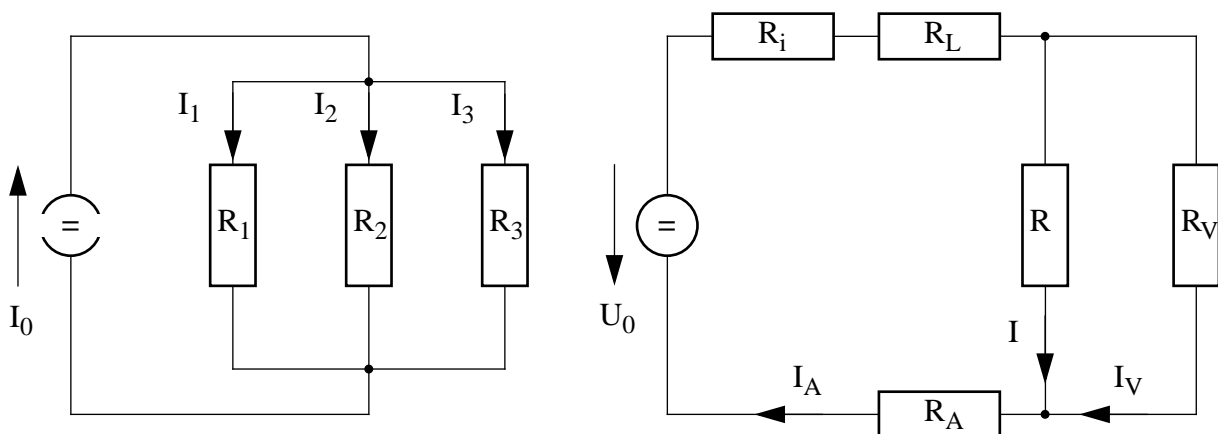


Bild 1.6: Zur Kirchhoffschen Knotenregel

1.3.5 Die magnetische Feldstärke

Während ein elektrisches Feld bereits von einer ruhenden Ladung ausgeht, entsteht ein „magnetisches Feld“ erst in der Umgebung bewegter Ladungen bzw. stromdurchflossener elektrischer Leiter. Beim magnetischen Feld liegt, wie beim elektrischen, eine doppelte Wechselwirkung vor:

- die Kraftwirkung, die ein vorhandenes Magnetfeld auf eine bewegte Ladung ausübt, und
- eine bewegte Ladung als Ursache eines Magnetfeldes.

Das magnetische Feld einer bewegten Ladung vermag auf eine andere bewegte Ladung Q eine Kraft F_m auszuüben, die bekanntlich eine gerichtete Größe („Vektor“) ist; die „magnetische Induktion“ B ist daher ebenfalls ein Vektor:

$$\vec{F}_m = Q \cdot (\vec{v} \times \vec{B}) \qquad (1.11)$$

Der Vektor der wirkenden Kraft F_m steht senkrecht auf den beiden Vektoren der magnetischen Induktion B und der Geschwindigkeit v , mit der sich die Ladung Q durch das Magnetfeld bewegt. Ein stromdurchflossener Leiter wird stets von magnetischen Feldlinien umhüllt, die in sich geschlossen sind (während elektrische Feldlinien auf Ladungen beginnen und enden); ihr Drehsinn ist per definitionem rechtsdrehend zur Stromflußrichtung. Damit lassen sich auch Anziehung und Abstoßung wie in Bild 1.7 veranschaulichen.

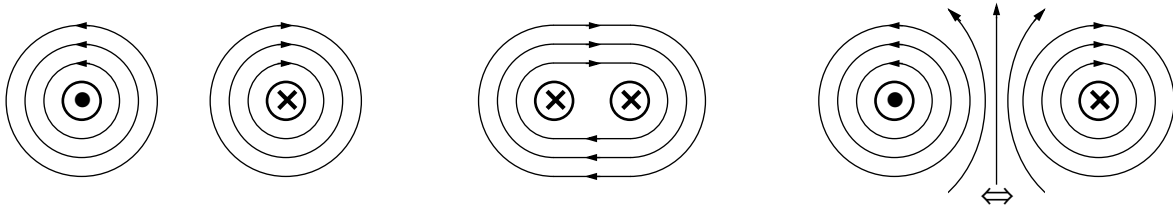


Bild 1.7: Magnetische Feldlinien bewegter Ladungen; • Stromfluß zum Betrachter, x weg vom Betrachter; links: einzelne Stromflüsse, mitte: Anziehung, rechts: Abstoßung.

In der Umgebung eines Stromflusses (und damit auch eines oder mehrerer stromdurchflossener Leiter) gilt:

$$\oint \vec{B} d\vec{A} = 0 \quad (1.12)$$

was besagt, daß im Magnetfeld in eine geschlossene Hüllfläche A ebenso viele Vektoren der magnetischen Induktion B eintreten wie aus ihr wieder austreten, d.h. das Magnetfeld ist quellenfrei. Mathematisch wird ein solches Integral bekanntlich als „Fluß“ bezeichnet, so daß man auch von einer „magnetischen Flußdichte“ B sprechen kann. Sie ist nach Gl.(1.11) die Wirkung des Magnetfeldes; dessen Ursache ist jedoch eine „magnetische Feldstärke“ H . Beide Größen sind zueinander proportional:

$$\vec{B} = \mu \cdot \vec{H} = \mu_r \mu_0 \cdot \vec{H} \quad \mu_0 = 4\pi \cdot 10^{-7} \text{Vs/Am} \quad (1.13)$$

Proportionalitätsfaktor ist die „Permeabilitätskonstante“ μ , wobei μ_0 für das Vakuum gilt, während die relative Permeabilitätszahl μ_r vom Material abhängt, in dem das Magnetfeld verläuft. Bei gleicher Feldstärke H ergeben sich je nach Material stark unterschiedliche Flüsse B und damit Kraftwirkungen des Magnetfeldes:

- $\mu_r = 1$: Luft; $\mu_r \gg 1$: ferromagnetisch (z.B. Eisen, Nickel, Kobalt);
 $\mu_r \leq 1$: diamagnetisch (z.B. Kupfer); $\mu_r \geq 1$: paramagnetisch (z.B. Aluminium).

Integriert man die magnetische Feldstärke H entlang eines geschlossenen Umlaufs, so ergibt sich der sie verursachende Strom i , der durch die umschlossene Fläche fließt:

$$\oint \vec{H} d\vec{s} = i \quad (1.14)$$

Wie man sieht, ist dieser Zusammenhang unabhängig von der Permeabilität μ , d.h. vom Material. Bei einem geraden, linienförmigen Stromleiter ist es am einfachsten, für die Berechnung einen geschlossenen Umlauf konstanter magnetischer Feldstärke zu wählen, d.h. einen konzentrischen Kreis um den Leiter (vgl. dazu Bild 1.7).

1.4 Elektrischer Widerstand

Physikalisch kann man sich das Wirken des elektrischen Widerstandes, der beim Anlegen einer elektrischen Spannung an einen metallischen Leiter die Stromstärke festlegt, anhand des Atommodells erklären. Beim Erstarren aus der Schmelze fügen sich die Metallatome zu einem Kristallgitter zusammen, wobei jedes Metallatom ein Elektron abgibt und als positiv geladenes Ion zurückbleibt. Die Elektronen bewegen sich quasi-frei durch das Kristallgitter, unter dem alleinigen Einfluß der Temperatur völlig zufällig und ungerichtet. Beim Anlegen einer Spannung überlagert sich dieser Bewegung die bereits erwähnte Vorzugsrichtung der Elektronen vom Minus- zum Pluspol der Spannungsquelle. Bei dieser Bewegung können die Elektronen mit den Ionen des Kristallgitters kollidieren oder an diesen gestreut werden, was sich nach außen hin meßtechnisch als elektrischer Widerstand äußert, den der Leiter dem Stromfluß entgegensetzt. Beim Vorliegen hoher Stromdichten kann es sogar zu einer Bewegung der Metallionen kommen, d.h. zu einem Transport von Leittermaterial, wobei der metallische Leiter schließlich unterbrochen wird.

1.4.1 Das Ohmsche Gesetz

Meßtechnisch beobachtet man bei gleichförmiger („stationärer“) elektrischer Strömung durch einen metallischen Leiter einen linearen Zusammenhang zwischen der Spannung U und dem Strom I :



Der Proportionalitätsfaktor wird als (elektrischer) „Widerstand“ bezeichnet; man verwendet dafür das Zeichen R :

$$U = R \cdot I \qquad [R] = V/A = \Omega \qquad (1.15)$$

Die Dimension $[R]$ des elektrischen Widerstandes heißt Ω (sprich: Ohm), benannt nach dem deutschen Physiker *Georg Simon Ohm* (1789-1854), der den linearen Zusammenhang zwischen Spannung, Strom und Widerstand als erster erkannt hat. Er formulierte das nach ihm benannte und in Gl.(1.15) wiedergegebene „Ohmsche Gesetz“, das in der Elektrotechnik eine zentrale Rolle spielt. Der Kehrwert des elektrischen Widerstandes wird als (elektrischer) „Leitwert“ mit dem Zeichen G bezeichnet:

$$G = 1/R \qquad [G] = A/V = S \qquad (1.16)$$

Die Dimension $[G]$ des elektrischen Leitwerts heißt S(iemens), benannt nach *Werner von Siemens* (1816-1892), dem Begründer der Elektrotechnik.

Wird das Ohmsche Gesetz für $I \Leftrightarrow y$ und $U \Leftrightarrow x$ durch eine lineare Funktion $y = f(x)$ dargestellt, so kann deren Steigung bekanntlich wie folgt bestimmt werden:

$$y = f(x) = m \cdot x \qquad \frac{dy}{dx} = m = \frac{dI}{dU} = 1/R = G$$

Das heißt, die Spannung U ist die unabhängige, der Strom I die abhängige Variable, und die Steigung der Geraden ist proportional zum Leitwert G bzw. umgekehrt proportional zum Widerstand R , wie Bild 1.8 zeigt.

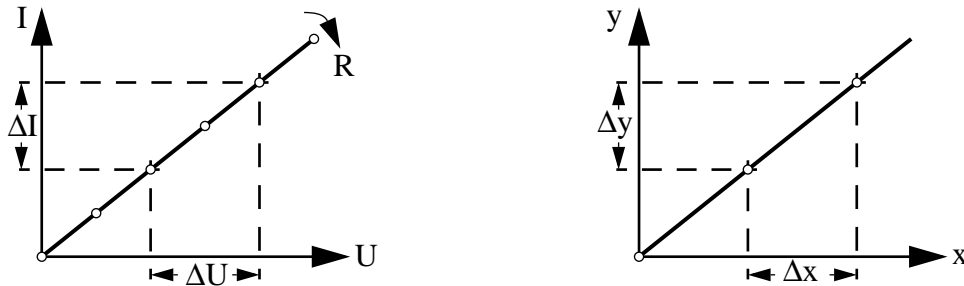


Bild 1.8: Zum Ohmschen Gesetz („Widerstandskennlinie“)

1.5 Energie im Stromkreis

1.5.1 Die elektrische Arbeit

Der Begriff der „elektrischen Arbeit“ wurde bereits bei der Einführung des Spannungsbegriffs verwendet, ohne ihn näher zu erläutern. Die Voraussetzung und das Maß für die Verrichtung von Arbeit ist Energie. Die physikalischen Größen Energie und Arbeit sind deshalb gleichwertig („äquivalent“); sie werden mit dem Zeichen W bezeichnet. Nach Gl.(1.6) und Gl.(1.8) ergibt sich für die bei der gerichteten Bewegung von Ladung umgesetzte Energie:

$$W = U \cdot Q = U \cdot I \cdot t \quad [W] = [U][I][t] = \text{VA s} = \text{J} \quad (1.17)$$

Dabei ist U die elektrische Spannung, Q die elektrische Ladung, I die elektrische Stromstärke und t die Zeit. Mit Gl.(1.17) läßt sich die Dimensionsgleichung für die elektrische Arbeit aufstellen: Die Dimension $[W]$ der elektrischen Arbeit heißt J(oule), benannt nach dem englischen Physiker *James Prescott Joule* (1818-1889); er fand das „Joulesche Gesetz“ über die in einem stromdurchflossenen Widerstand umgesetzte Wärmemenge. Man verwechsle aber nicht die Dimension $[W] = \text{J(oule)}$ mit der Stromdichte J , deren Dimension $[J] = \text{A/mm}^2$ ist.

Die in einem Widerstand in Wärme umgesetzte elektrische Arbeit läßt sich aus Gl.(1.17) in Verbindung mit dem Ohmschen Gesetz Gl.(1.15) herleiten:

$$W = \frac{U^2}{R} \cdot t = I^2 \cdot R \cdot t \quad (1.18)$$

1.5.2 Die elektrische Leistung

Für die „elektrische Leistung“ verwendet man das Zeichen P ; sie ist als elektrische Arbeit pro Zeiteinheit definiert:

$$P = \frac{W}{t} = \frac{U \cdot I \cdot t}{t} = U \cdot I \quad [P] = [U][I] = \text{VA} = \text{W} \quad (1.19)$$

Dabei ist W die elektrische Arbeit, U die Spannung, I die Stromstärke und t die Zeit. Die Dimension $[P]$ der elektrischen Leistung heißt $W(\text{att})$, benannt nach dem englischen Erfinder *James Watt* (1736-1819), dem Erbauer der ersten Dampfmaschine. Man verwechsle jedoch nicht die Dimension $[P] = W(\text{att})$ mit der elektrischen Arbeit W , deren Dimension $[W] = J$ ist.

In Verbindung mit dem Ohmschen Gesetz Gl.(1.15) ergibt sich mit Gl.(1.19) für die in einem elektrischen Widerstand R umgesetzte elektrische Leistung:

$$P = \frac{U^2}{R} = I^2 \cdot R \quad (1.20)$$

Für elektrische Bauelemente ist der maximal zulässige Wert der umgesetzten Leistung eine wichtige Kenngröße bei der Dimensionierung elektrischer Schaltungen, da bei Überschreiten des zulässigen Maximalwertes das Bauelement überhitzt und zerstört werden kann. Deshalb ist beim Bau elektrischer Geräte auch sorgfältig auf eine Abfuhr der erzeugten Wärme durch Kühlung zu achten, sei es durch selbsttätige Konvektion der umgebenden Luft, sei es explizit durch Ventilation oder Kühlmittelfluß.

Man pflegt komplexe elektrische Netzwerke aus Gründen der Übersichtlichkeit in Blöcke zu gliedern, die jeweils mehrere Bauelemente enthalten. Diese Blöcke können wiederum zu noch umfassenderen Einheiten zusammengefaßt werden. Diese „aufbauende“ Vorgehensweise wird im englischen Sprachraum mit „bottom-up“ bezeichnet. Umgekehrt kann man auch ein System Schritt für Schritt in Einheiten, Blöcke und Bauelemente „zergliedern“, was in der englischsprachigen Literatur als „top-down“ bezeichnet wird. Man erhält in beiden Fällen eine hierarchische Gliederung komplexer Systeme in Komponenten, die miteinander kommunizieren. Je nach Anzahl der äußeren Anschlüsse lassen sich zwei Klassen von Komponenten unterscheiden (Bild 2.1):

- „Zweipole“: Einfache Bauelemente, die nur zwei Anschlußklemmen besitzen, wie Spannungs- oder Stromquellen, Widerstände, Kondensatoren und Spulen, lassen sich als Zweipole darstellen.
- „Vierpole“: Komplexere Einheiten, wie Verstärker oder Frequenzfilter, besitzen zwei Eingangs- und zwei Ausgangsklemmen; sie lassen sich als Vierpole darstellen.

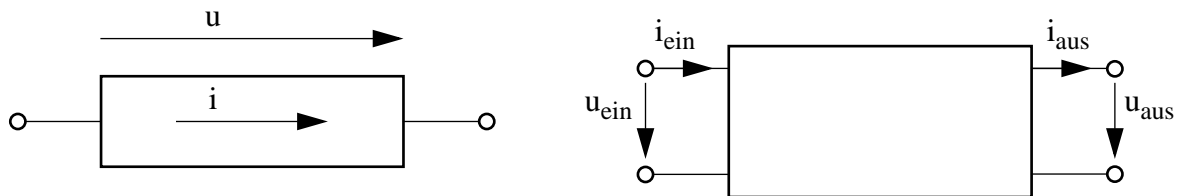


Bild 2.1: Zweipol und Vierpol

2.1 Darstellungen mit Zweipolen

2.1.1 Elektrische Stromkreise

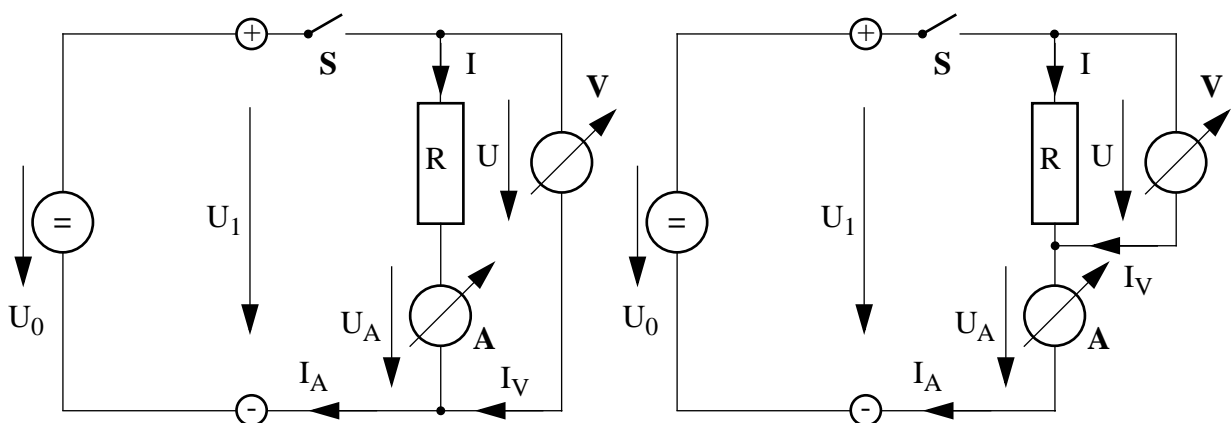


Bild 2.2: Meßschaltungen; A: Amperemeter, V: Voltmeter
links: stromrichtig, rechts: spannungsrichtig

Das technische Zusammenwirken der in Kapitel 1 vorgestellten Grundgrößen Spannung, Strom und Widerstand läßt sich durch „Stromkreise“ darstellen. Im einfachsten Fall besteht ein Stromkreis aus einer Spannungsquelle U_0 , einem Schalter S, einem Widerstand R als Verbraucher und den

Verbindungsleitungen (Bild 2.2). Die Ladungstrennung, die in der Spannungsquelle erfolgt, führt nach dem Schließen des Schalters zu einem Stromfluß durch den nunmehr geschlossenen Stromkreis, wobei die von der Quelle gelieferte elektrische Energie im Verbraucher in eine andere Energieform (z.B. Licht, Bewegung und/oder Wärme) umgesetzt wird. Die (negativ geladenen) Elektronen fließen vom Minus- zum Pluspol der Spannungsquelle, der technische Stromfluß erfolgt per definitionem in umgekehrter Richtung, wie es durch die Richtungen der Strom- und Spannungspfeile zum Ausdruck kommt. Die Werte der Ströme und Spannungen ergeben sich bei dieser Wahl der Pfeilrichtung zu > 0 .

Am Verbraucher R kann der Spannungsabfall U durch einen Spannungsmesser V („Voltmeter“), die Stromstärke I durch einen Strommesser A („Amperemeter“) gemessen werden, wie Bild 2.2 zeigt. Die Meßgeräte benötigen zur Anzeige eine wenn auch geringe Energie, was durch ihre Innenwiderstände R_V bzw. R_A dargestellt werden kann (Bild 2.3). Man unterscheidet eine „stromrichtige“ und eine „spannungsrichtige“ Meßschaltung: Im stromrichtigen Fall mißt das Voltmeter nicht nur die Spannung U am Verbraucher R , sondern auch den Spannungsabfall U_A am Innenwiderstand R_A des Amperemeters. Im spannungsrichtigen Fall mißt das Amperemeter nicht nur den Strom I durch den Verbraucher R , sondern auch den Strom I_V durch das Voltmeter mit dem Innenwiderstand R_V . Man wird durch geeignete Dimensionierung der Innenwiderstände ($R_A \ll R \ll R_V$) dafür sorgen, dass die prinzipiell nicht zu vermeidenden Meßfehler vernachlässigbar klein bleiben.

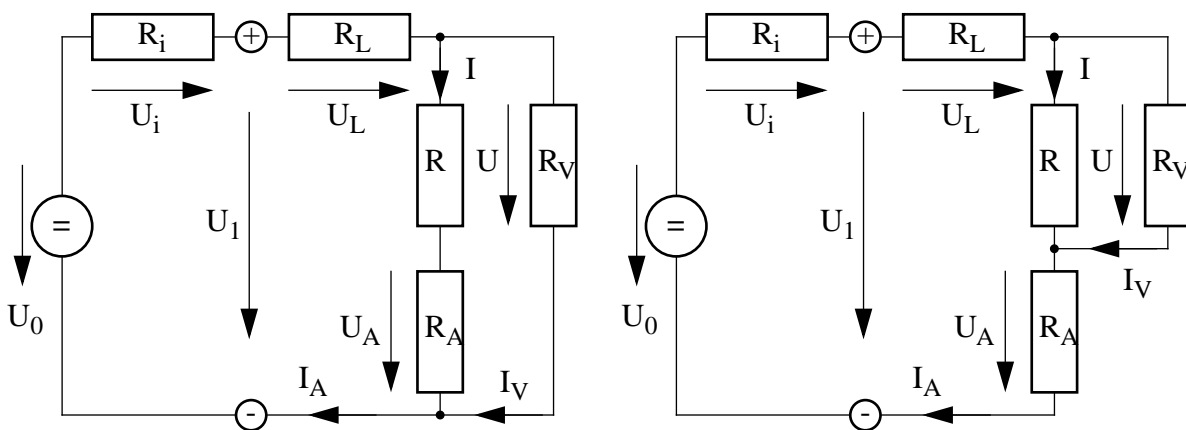


Bild 2.3: Einfache elektrische Stromkreise (vgl. mit Bild 2.2)

Die Ströme und Spannungen lassen sich anhand Bild 2.3 mittels der Kirchhoffschen Regeln und des Ohmschen Gesetzes ermitteln. Die Spannungsquelle weist zwar eine fest eingepreßte Spannung U_0 auf; die im realen Fall von ihr gelieferte Spannung U_1 nimmt jedoch mit zunehmendem Strom I_A ab, was durch einen Innenwiderstand R_i dargestellt werden kann. Der Widerstand der Zuleitungen wird in R_L zusammengefaßt. Im Idealfall wird ein Innenwiderstand $R_i = 0$, ein Zuleitungswiderstand $R_L = 0$, der Innenwiderstand des Strommessers ebenfalls zu $R_A = 0$ und der Innenwiderstand des Spannungsmessers zu $R_V \rightarrow \infty$ angenommen.

2.1.2 Elektrische Zweipole

Die größte Aufteilung eines elektrischen Stromkreises erfolgt in zwei Zweipole:

- einen Erzeugerzweig, der mindestens eine Energie erzeugende Spannungs- oder Stromquelle enthält,
- und einen Verbraucherzweig, der ein oder mehrere Energie verbrauchende Bauelemente enthält, wie Widerstände, Kondensatoren, Spulen und/oder Dioden.

Der tatsächliche innere Aufbau dieser Zweipole braucht nicht bekannt zu sein. Man kann vielmehr durch geeignete Messungen an den Anschlußklemmen der beteiligten Zweipole zu einer sogenannten *Ersatzschaltung* kommen, die aus möglichst wenigen und möglichst aus linearen Bauelementen zusammengesetzt ist und dasselbe - oder näherungsweise dasselbe - Verhalten zeigt. In der detailliertesten Aufteilung eines elektrischen Netzwerks enthalten alle Zweipole jeweils nur ein einziges Bauelement.

Passive Bauelemente der Elektrotechnik sind Verbraucherzweipole aufgrund ihres Aufbaus und ihrer Wirkungsweise. Sie verbrauchen elektrische Energie; genauer: sie wandeln sie in Wärmeenergie um. Man unterscheidet je nach funktionellem Zusammenhang zwischen durchfließendem Strom und angelegter Spannung:

- Ideale Zweipole: Widerstand R , Kapazität C , Induktivität L .
- Technische Zweipole, wie Widerstände, Kondensatoren und Spulen, d.h. nicht-ideale Bauelemente, deren Verhalten durch ein entsprechendes Ersatzschaltbild aus idealen Zweipolen näherungsweise dargestellt werden kann.
- Nichtlineare Zweipole, wie z.B. Gleichrichterioden, wobei der nichtlineare Zusammenhang zwischen Strom und Spannung ebenfalls durch Ersatzschaltbilder aus idealen Zweipolen näherungsweise nachgebildet werden kann.

2.2 Resistive Zweipole: Widerstände

2.2.1 Idealer elektrischer Widerstand

Für einen idealen Widerstand mit dem Widerstandswert R besteht nach dem Ohmschen Gesetz bekanntlich ein linearer Zusammenhang zwischen der angelegten Spannung u_R und dem Strom i_R :

$$u_R = R \cdot i_R$$

$$i_R = I_R \cdot \sin \omega t$$

$$u_R = R \cdot I_R \cdot \sin \omega t \quad (2.1)$$

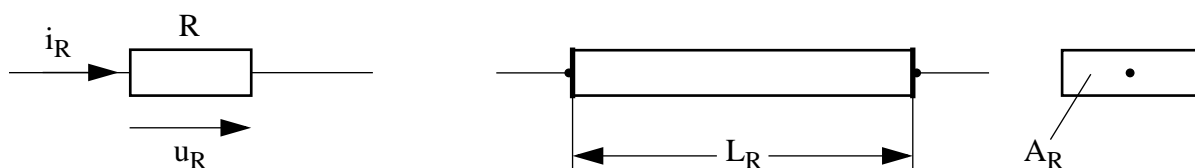


Bild 2.4: Schaltungssymbol und Bauprinzip eines Widerstandes

Mit Gl.(2.1) kann man zeigen, dass beim Betrieb mit sinusförmigem Wechselstrom der Strom i_R und die Spannung u_R zeitlich „phasengleich“ verlaufen, d.h. beim Widerstand besteht keine Phasenverschiebung φ zwischen Strom und Spannung.

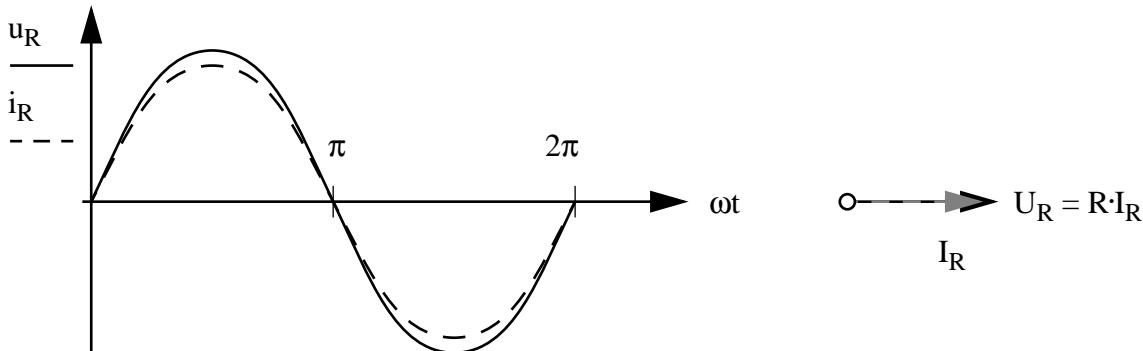


Bild 2.5: Phasenbild eines Widerstandes im Wechselstrombetrieb

Grenzfälle des Widerstandes R :

- Ein „Kurzschluß“ $R = 0$ liegt vor, wenn selbst bei beliebig hohem Stromfluß kein Spannungsabfall entsteht.
- Ein „Leerlauf“ $R \rightarrow \infty$ liegt vor, wenn auch bei beliebig hoher Spannung kein Strom fließt.

Der elektrische Widerstand R eines Leiters oder Bauelements ist proportional zur Länge L_R und umgekehrt proportional zur Querschnittsfläche A_R der Widerstandsbahn (Bild 2.4). Der Proportionalitätsfaktor ρ wird als „spezifischer Widerstand“ bezeichnet:

$$R = \rho \cdot \frac{L_R}{A_R} = \frac{1}{\sigma} \cdot \frac{L_R}{A_R} \quad [\rho] = [R] \cdot \frac{[A_R]}{[L_R]} = \Omega \cdot \frac{\text{mm}^2}{\text{m}} \quad (2.2)$$

Den Kehrwert des spezifischen Widerstandes bezeichnet man als „spezifische Leitfähigkeit“ σ :

$$\sigma = n \cdot e \cdot \mu \quad (2.3)$$

Sie ist proportional zur Elektronendichte n im betreffenden Material, d.h. zur Anzahl der Elektronen pro Volumeneinheit, und zur Beweglichkeit μ der Elektronen im Kristallgitter; Proportionalitätsfaktor ist die Elementarladung e . Anhand von Gl.(2.2) kann die Dimension $[\rho]$ des spezifischen Widerstands bestimmt werden. Er ist eine Materialkenngröße, die für das verwendete Leiter- bzw. Widerstandsmaterial charakteristisch ist. Für einige technisch wichtige Leitermaterialien gilt:

- Silber: $\rho_{\text{Ag}} = 0,016 \quad \Omega\text{mm}^2/\text{m}; \quad n_{\text{Ag}} = 5,87 \cdot 10^{22} \text{ Elektronen}/\text{cm}^3$
- Kupfer: $\rho_{\text{Cu}} = 0,0178 \quad \Omega\text{mm}^2/\text{m}; \quad n_{\text{Cu}} = 8,47 \cdot 10^{22} \text{ Elektronen}/\text{cm}^3$
- Aluminium: $\rho_{\text{Al}} = 0,027 \quad \Omega\text{mm}^2/\text{m};$
- Kohle: $\rho_{\text{C}} = 40 \dots 100 \Omega\text{mm}^2/\text{m};$
- Silizium: $\rho_{\text{Si}} = 2,3 \cdot 10^9 \quad \Omega\text{mm}^2/\text{m}; \quad n_{\text{Si}} = 10^{10} \dots 10^{20} \text{ Elektronen}/\text{cm}^3$

Für flächenhafte Widerstandsschichten, wie sie vor allem in Dünnschicht- und hochintegrierten Halbleiterschaltungen der Mikroelektronik vorkommen, wurde der „Schichtwiderstand“ R_S eingeführt, der sich auf die Dicke d der Widerstandsschicht bezieht:

$$R = \frac{\rho}{d} \cdot \frac{L_R}{B_R} = R_S \cdot \frac{L_R}{B_R} \quad \text{wobei } A_R = d \cdot B_R \quad (2.4)$$

Da der Schichtwiderstand R_S durch die Technologie vorgegeben ist, hängt der Widerstandswert R nur noch von der Auslegung („Layout“), das heißt vom Längen-/Breitenverhältnis der Widerstandsbahn ab. Typische Zahlenwerte sind $R_S = 10 \dots 500 \, \Omega/\square$; sie entsprechen dem Widerstandswert eines Quadrats in der betreffenden Technologie.

2.2.2 Technische Widerstände

Technische Widerstände dienen u.a. zur Strombegrenzung und Spannungsteilung bzw. zur Erzeugung von Potentialverschiebungen. Sie sind durch Nennwert, Toleranz, Belastbarkeit, Temperaturabhängigkeit und nicht zuletzt durch ihren Aufbau gekennzeichnet. Sie werden durch einen idealen („ohmschen“) Widerstand nur unvollkommen beschrieben. Man unterscheidet im wesentlichen folgende Bauformen (Bild 2.6):

- Schichtwiderstände aus massiver Keramik oder einem Glaskörper, auf denen eine dünne Metall- oder Kohleschicht aufgebracht ist. Metallschichtwiderstände sind besonders zuverlässig mit einer zulässigen Verlustleistung von bis zu 10 W.
- Drahtwiderstände aus einem Keramikkörper, der eine Wicklung aus Widerstandsdraht trägt. Sie eignen sich für höhere Verlustleistungen bis ca. 25 W. Sehr zweckmäßig sind Mehrfachwiderstände in einem Keramikkörper, die sich zu unterschiedlichen Widerstandswerten kombinieren lassen.
- Widerstandsmodule, die mehrere Widerstände auf einem Keramikträger in Dünnschicht- oder Dickschichttechnik enthalten. Meist werden solche Module bei entsprechender Stückzahl nach Anwendervorgabe hergestellt.
- Massewiderstände aus einer homogenen Mischung von Widerstandsmaterial und organischem Bindemittel, die zu einem zylindrischen Körper geformt wird, in den die Anschlußdrähte eingebettet sind und der mit isolierendem Kunststoff umhüllt ist. Die zulässige Verlustleistung beträgt bis zu 2 W.
- Potentiometer und Trimmer, d.h. verstellbare Widerstände mit drei Anschlüssen am Anfang und Ende der Widerstandsbahn sowie an einem Schleifkontakt, der sich mechanisch einstellen läßt.

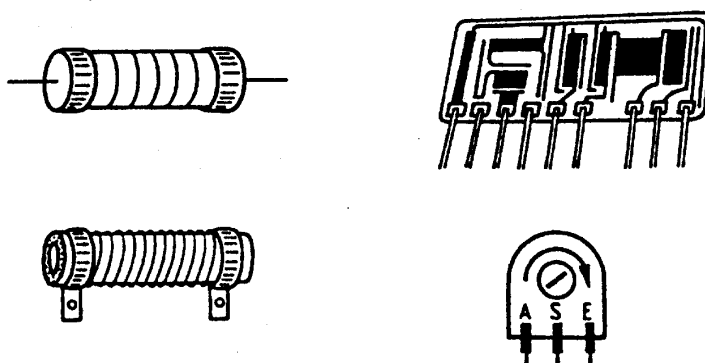


Bild 2.6: Bauformen technischer Widerstände

Da Schichtwiderstände zur Befestigung der äußeren Anschlußdrähte mit Metallkappen versehen werden, weisen sie eine parasitäre Kapazität auf, die parallel zum eigentlichen Widerstandswert liegt. Beim Einsatz von Drahtwiderständen ist auf deren parasitäre Induktivität zu achten, die durch das Aufwickeln des Widerstandsdrahtes entsteht und in Reihe zum eigentlichen Widerstand liegt. Abhängig von der Betriebsfrequenz der technischen Widerstände müssen daher ihre kapazitiven und/oder induktiven Anteile beachtet und ggf. in einem Ersatzschaltbild erfaßt werden (Bild 2.7).

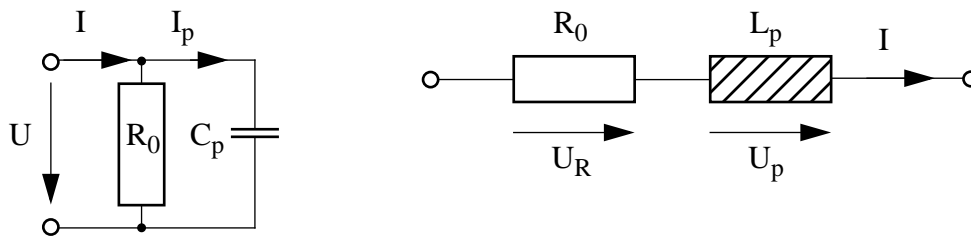


Bild 2.7: Ersatzschaltbilder technischer Widerstände
links: Schichtwiderstand, rechts: Drahtwiderstand

Kennzeichnung. Aus Gründen der Wirtschaftlichkeit werden nicht beliebige Widerstandswerte hergestellt; sie sind vielmehr nach DIN 41426 genormt. Die Darstellung des Widerstandswerts und des Toleranzbereichs eines technischen Widerstands erfolgt durch einen genormten Farbcode, d.h. Ringe, Striche oder Punkte, die auf der Oberfläche des Bauelements angebracht sind, oder bei genügend großen Bauformen durch einen Zahlenaufdruck.

2.2.3 Schaltungen mit idealen Widerständen

Durch Widerstände, die wie in Bild 2.8 (links) in Reihe geschaltet sind, fließt derselbe Strom I. Nach dem Ohmschen Gesetz ($U = R \cdot I$) und der Kirchhoffschen Maschenregel („In jeder Masche ist die Summe der Spannungen gleich Null“) gilt dann für den Gesamtwiderstand der elektrischen Reihenschaltung:

$$R_{r, \text{ges}} = \frac{U_0}{I} = \frac{U_1 + U_2 + U_3}{I} = R_1 + R_2 + R_3 \quad (2.5)$$

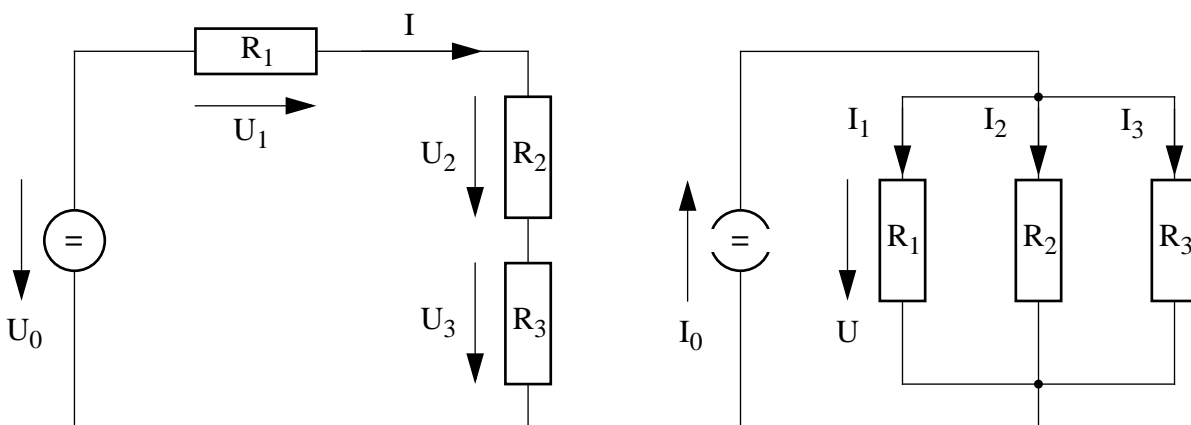


Bild 2.8: Reihenschaltung (links), Parallelschaltung (rechts) von idealen Widerständen

An Widerständen, die wie in Bild 2.8 (rechts) parallel geschaltet sind, liegt dieselbe Spannung U . Nach dem Ohmschen Gesetz gilt für den Gesamtleitwert der elektrischen Parallelschaltung:

$$G_{p, ges} = \frac{I_0}{U} = \frac{I_1 + I_2 + I_3}{U} = G_1 + G_2 + G_3 = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} \quad (2.6)$$

2.3 Induktive Zweipole: Spulen

2.3.1 Ideale elektrische Induktivität

Für eine ideale Spule mit der Induktivität L gilt folgender Zusammenhang zwischen der angelegten Spannung u_L und dem Strom i_L :

$$u_L = L \cdot \frac{di_L}{dt} \quad i_L = -I_L \cdot \cos \omega t$$

$$u_L = \omega L \cdot I_L \cdot \sin \omega t \quad (2.7)$$

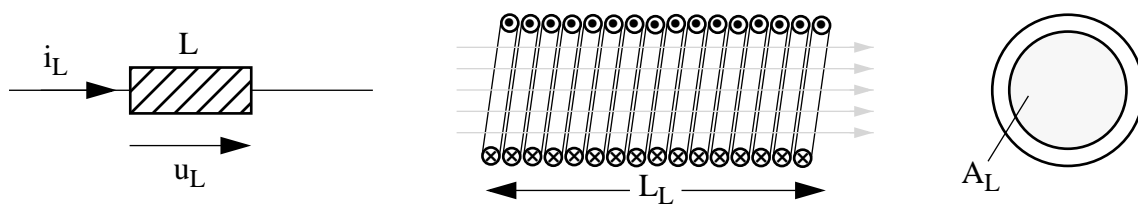


Bild 2.9: Schaltungssymbol und Bauprinzip einer Spule

Gl. (2.7) besagt, dass nur *zeitliche Änderungen* des magnetischen Feldes, das aufgrund des Stromes i_L in der Spule herrscht, zu einem Spannungsabfall u_L an der Spule führen. Ein formaler Vergleich mit dem Ohmschen Gesetz in der Form von Gl.(2.1) zeigt, dass die Größe $\omega L = X_L$ als Widerstand („Blindwiderstand“) aufgefaßt werden kann. Liegt an der Spule eine sinusförmige Wechselspannung u_L , so hinkt der Strom i_L der Spannung um $\pi/2$ nach, d.h. bei der Spule besteht zwischen Spannung und Strom eine Phasenverschiebung $\varphi = -90^\circ$.

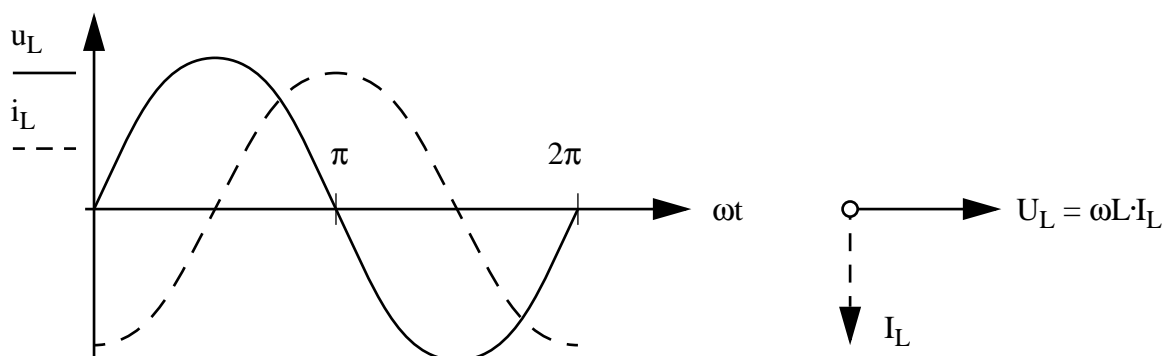


Bild 2.10: Phasenbild einer Spule im Wechselstrombetrieb

2.3.2 Technische Spulen

Spulen finden ihre hauptsächliche Anwendung zur Unterdrückung von Hochfrequenzeinwirkungen auf die Gleichstromversorgung oder in Verbindung mit Kondensatoren in Schwingkreisen. In Schaltungen der Mikroelektronik werden sie wegen ihrer relativ großen Abmessungen kaum verwendet. Man muss zwei Gruppen unterschiedlicher Bauformen unterscheiden: Luftspulen, bei denen gut leitender Draht auf einen isolierenden Hohlkörper gewickelt wird, und Spulen mit Magnetkern, der zur Führung und Verstärkung des magnetischen Feldes dient.

Für die Induktivität einer geraden, im Vergleich zu ihrem Durchmesser hinreichend langen Luftspule gilt:

$$L_0 = \mu_0 \cdot N^2 \cdot \frac{A_L}{L_L} \quad [L] = \frac{V \cdot s}{A} = H \quad (2.8)$$

Dabei ist $\mu_0 = 4\pi \cdot 10^{-7} \text{ Vs/Am}$ die „Permeabilitätskonstante“ des Vakuums (sie gilt näherungsweise auch für Luft), N die Windungszahl der Luftspule, A_L der wicklungsfreie Querschnitt und L_L ihre Länge (Bild 2.9). Die Dimension $[L]$ der Induktivität heißt H(enry), benannt nach dem amerikanischen Physiker *Joseph Henry* (1797 - 1878), dem Entdecker der Induktionserscheinungen. Die Einheit 1 H ist für praktische Anwendungen zu groß, weshalb man folgende Größenordnungen festlegt: $1 \text{ nH} = 10^{-9} \text{ H}$; $1 \text{ }\mu\text{H} = 10^{-6} \text{ H}$; $1 \text{ mH} = 10^{-3} \text{ H}$

Zur Verbesserung der magnetischen Kopplung oder um höhere Induktivitätswerte zu erreichen, werden Spulen mit einem Kern aus magnetischem Werkstoff versehen. Für den Wert der Induktivität gilt dann:

$$L = \mu_r \cdot \mu_0 \cdot N^2 \cdot \frac{A_k}{L_k} \quad (2.9)$$

wobei μ_r die (dimensionslose) relative Permeabilitätszahl des magnetischen Werkstoffes des Spulenkerns ist, N die Windungszahl, A_k die Querschnittsfläche des Kerns und L_k die mittlere Feldlinienlänge im Kern. Bei den Magnetkernen unterscheidet man zwischen Kernblechen, die für Transformatoren verwendet werden, und Ferritkernen für nachrichtentechnische Anwendungen.

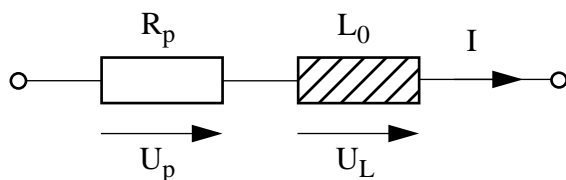


Bild 2.11: Ersatzschaltbild technischer Spulen

Für den Einsatz von Spulen in elektronischen Schaltungen ist die Kenntnis ihrer technischen Eigenschaften wichtig. Es gelten folgende Kenngrößen: Spulengüte (Blind- zu Wirkwiderstand), effektive Permeabilität (evtl. durch einen Luftspalt im Kern verringert), mittlerer Wicklungsquerschnitt (der für die Spulendrähte zur Verfügung steht) und der Drahtdurchmesser. Letzterer bestimmt die zulässige Stromdichte, da die in der Spule erzeugte Wärme abgeführt werden muss. Ferner ist der Drahtwiderstand des Wicklungsdrahtes nicht zu vernachlässigen; er ist, wie in Bild 2.11, in einem Ersatzschaltbild zu berücksichtigen.

2.3.3 Schaltungen mit idealen Induktivitäten

Analog zur Betrachtung für ohmsche Widerstände nach Gl.(2.5) gilt auch für den gesamten Blindwiderstand $X_{L,ges}$ mehrerer, wie in Bild 2.12 (links) in Reihe geschalteter Induktivitäten:

$$\begin{aligned} X_{L,ges} &= \omega L_{r,ges} = X_{L1} + X_{L2} + X_{L3} = \omega L_1 + \omega L_2 + \omega L_3 \\ L_{r,ges} &= L_1 + L_2 + L_3 \end{aligned} \quad (2.10)$$

Entsprechend kann man für die Parallelschaltung von Induktivitäten nach Bild 2.12 (rechts) nach der für ohmsche Widerstände geltenden Gleichung Gl.(2.6) verfahren, die auch für den Kehrwert $Y_L = 1/X_L = 1/\omega L$ von Blindwiderständen gilt.

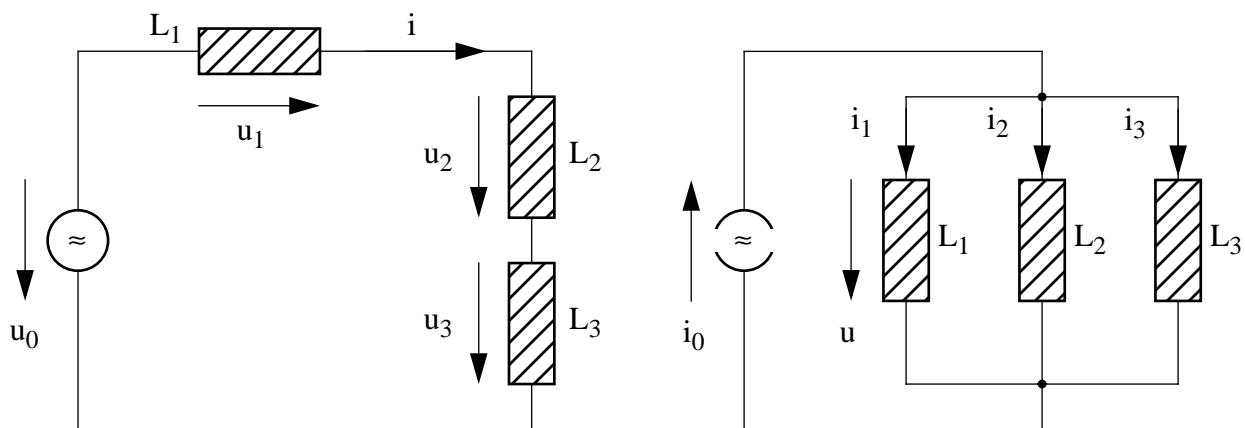


Bild 2.12: Reihenschaltung (links), Parallelschaltung (rechts) von idealen Induktivitäten

2.4 Kapazitive Zweipole: Kondensatoren

2.4.1 Die elektrische Kapazität

Für einen idealen Kondensator mit der Kapazität C gilt folgender Zusammenhang zwischen der angelegten Spannung u_C und dem Strom i_C :

$$\begin{aligned} i_C &= C \cdot \frac{du_C}{dt} & u_C &= U_C \cdot \sin \omega t \\ i_C &= \omega C \cdot U_C \cdot \cos \omega t \end{aligned} \quad (2.11)$$

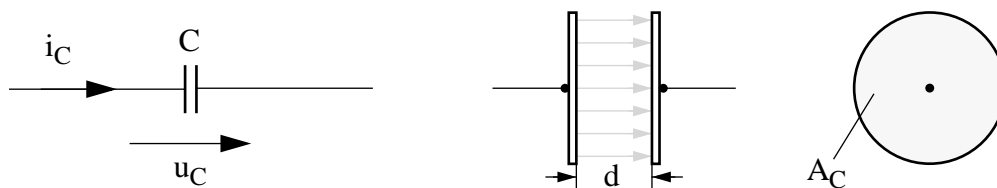


Bild 2.13: Schaltungssymbol und Bauprinzip eines Kondensators

Gl. (2.11) besagt, dass nur *zeitliche Änderungen* des elektrischen Feldes, das aufgrund der angelegten Spannung u_C im Kondensator herrscht, zu einem Stromfluß i_C durch den Kondensator führen.

Ein formaler Vergleich mit dem Ohmschen Gesetz in der Form von Gl.(2.1) zeigt, dass die Größe $\omega C = Y_C$ als Leitwert („Blindleitwert“) aufgefaßt werden kann. Liegt am Kondensator eine sinusförmige Wechselspannung u_C , so eilt der Strom i_C der Spannung um $\pi/2$ voraus, d.h. beim Kondensator besteht zwischen Spannung und Strom eine Phasenverschiebung $\varphi = 90^\circ$.

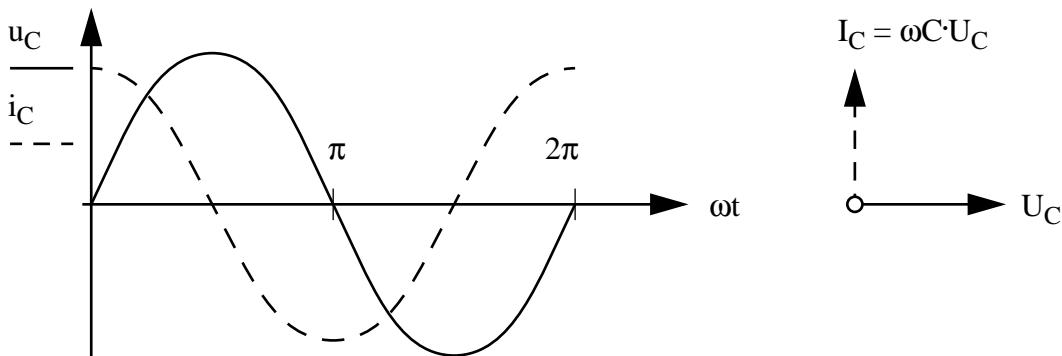


Bild 2.14: Phasenbild eines Kondensators im Wechselstrombetrieb

2.4.2 Technische Kondensatoren

Die hauptsächlichen Anwendungsbereiche von Kondensatoren sind die Trennung von Gleich- und Wechselstromkreisen, allgemein die Trennung hoher von tiefen Frequenzen als Hoch- oder Tiefpaß. Außerdem werden sie in Verbindung mit Spulen in Schwingkreisen verwendet.

Zwei metallisch leitende Flächen, die durch eine nichtleitende Schicht („Dielektrikum“) getrennt sind, bilden einen Kondensator mit der Kapazität:

$$C = \epsilon_r \cdot \epsilon_0 \cdot \frac{A_C}{d} \qquad [C] = \frac{A \cdot s}{V} = F \qquad (2.12)$$

Dabei ist $\epsilon_0 = 0,885 \cdot 10^{-11} \text{ As/Vm}$ die „Dielektrizitätskonstante“ des Vakuums (sie gilt näherungsweise auch für Luft), ϵ_r die (dimensionslose) relative Dielektrizitätszahl, A_C die Fläche und d der Abstand der Kondensatorplatten bzw. die Dicke des Dielektrikums (Bild 2.13). Die Dimension $[C]$ der Kapazität heißt F(arad), benannt nach dem englischen Naturwissenschaftler *Michael Faraday* (1791 - 1867), dem Entdecker der Grundgesetze der Elektrolyse. Für praktische Anwendungen ist die Einheit 1 F viel zu groß, weshalb man folgende Größenordnungen festlegt: $1 \text{ fF} = 10^{-15} \text{ F}$; $1 \text{ pF} = 10^{-12} \text{ F}$; $1 \text{ nF} = 10^{-9} \text{ F}$; $1 \mu\text{F} = 10^{-6} \text{ F}$; $1 \text{ mF} = 10^{-3} \text{ F}$

Man unterscheidet zahlreiche Bauformen, von denen nachfolgend nur einige genannt werden sollen:

- Eine kompakte Bauform ist der Keramikkondensator. Das Dielektrikum besteht aus Keramikmasse mit einer besonders hohen relativen Dielektrizitätszahl $\epsilon_r = 10 \dots 100$. Keramikkondensatoren eignen sich für hohe Frequenzen und hohe Betriebsspannungen; es können allerdings nur relativ kleine Kapazitätswerte von einigen nF erreicht werden.
- Relativ hohe Kapazitätswerte von einigen mF lassen sich nur mit Elektrolytkondensatoren herstellen. Diese bestehen aus einem aufgewickelten Aluminiumband, das die eine Kondensatorfläche bildet; der zweite Belag ergibt sich durch einen flüssigen Elektrolyten, während eine extrem dünne Oxidschicht des Aluminiums das Dielektrikum bildet. Elektrolytkondensatoren sind in der Regel gepolt, um die Oxidschicht auf der Anode zu gewährleisten.

- **Veränderbare Kondensatoren** werden als Plattenkondensatoren ausgeführt, wobei ein Plattensatz drehbar ist und kammförmig in einen festen Plattensatz eingreift. Dadurch kann die Fläche des elektrischen Feldes im Kondensator verändert werden. Das Dielektrikum ist Luft. Der Variationsbereich liegt bei relativ niedrigen Werten von 10 . . 400 pF. Sie werden zum Abgleich von Schwingkreisen eingesetzt.

Falls technische Kondensatoren einen kleinen Reststrom („Leckstrom“) führen, kann dieser wie in Bild 2.15 ersatzweise durch einen zum Kondensator parallelen Leitwert dargestellt werden.

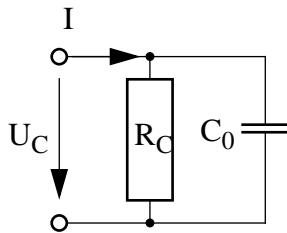


Bild 2.15: Ersatzschaltbild technischer Kondensatoren

Kennzeichnung. Für den Einsatz von Kondensatoren in elektronischen Schaltungen ist die Kenntnis ihrer technischen Eigenschaften wichtig. Es gelten folgende Kenngrößen: Nennkapazität, Frequenzabhängigkeit, Nennspannung, Spitzenspannung, Verlustfaktor (Blind- zu Wirkstrom) und Temperaturabhängigkeit. Die Kapazitätswerte und Toleranzbereiche sind nach DIN 41426 genormt. Die Kennzeichnung erfolgt durch Zahlen- und Buchstabenaufdruck auf dem Gehäuse oder, ähnlich wie bei Widerständen, durch einen (anderen) genormten Farbcode.

2.4.3 Schaltungen mit idealen Kapazitäten

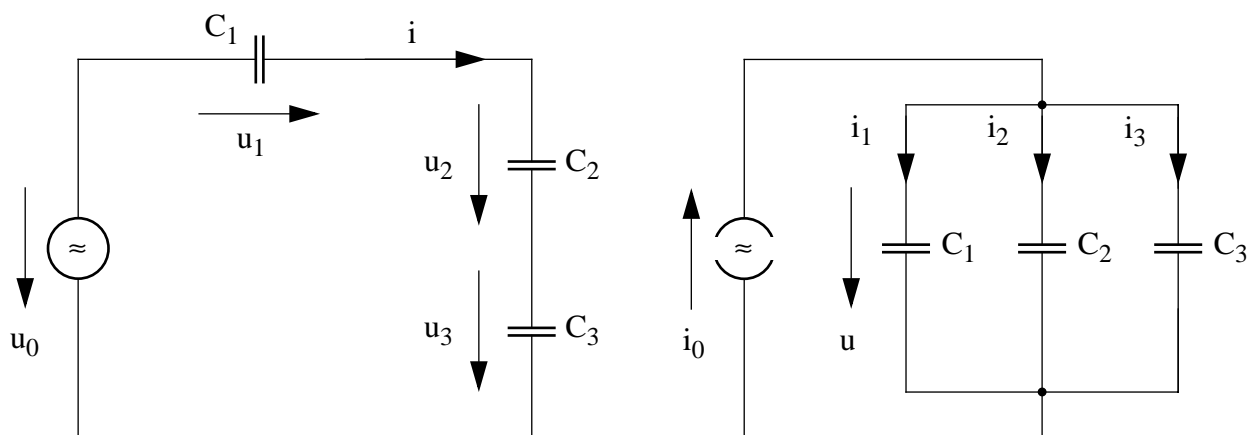


Bild 2.16: Reihenschaltung (links), Parallelschaltung (rechts) von idealen Kapazitäten

Analog zu der Betrachtung für ohmsche Widerstände nach Gl.(2.6) gilt auch für den gesamten Blindleitwert $Y_{C,ges}$ mehrerer, wie in Bild 2.16 (rechts) parallel geschalteter Kapazitäten:

$$Y_{Cp,ges} = \omega C_{p,ges} = Y_{C1} + Y_{C2} + Y_{C3} = \omega C_1 + \omega C_2 + \omega C_3$$

$$C_{p,ges} = C_1 + C_2 + C_3 \quad (2.13)$$

Entsprechend kann man für die Reihenschaltung von Kapazitäten nach Bild 2.16 (links) nach der

für ohmsche Widerstände geltenden Gleichung Gl.(2.5) vorgehen, die auch für den Kehrwert $X_C = 1/Y_C = 1/\omega C$ von Blindleitwerten gilt.

2.5 Nichtlineare Zweipole

2.5.1 Silizium-Dioden

Während für (ideale) Widerstände, Kondensatoren und Spulen ein linearer Zusammenhang zwischen Strom und Spannung besteht, gekennzeichnet durch Proportionalitätsfaktoren (Widerstandswert R , Kapazität C und Induktivität L), zeigt eine Gleichrichterdiode aufgrund anderer physikalischer Zusammenhänge ein nichtlineares Strom-Spannungsverhalten. Dabei ist zwischen der idealen Diodenkennlinie und der Strom-Spannungskennlinie realer Bauelemente zu unterscheiden.

Ideale Diode:
$$I_D = I_{S0} \cdot \left[\exp\left(\frac{U_D}{U_T}\right) - 1 \right] \quad (2.14)$$

Dabei ist I_{S0} der Sättigungsstrom im Sperrbereich, $U_T = kT/e$ die Temperaturspannung (sie beträgt bei Zimmertemperatur etwa 26 mV).

Reale Silizium-Diode:
$$I_D = I_S \cdot \left[\exp\left(\frac{U_D}{a \cdot U_T}\right) - 1 \right] \quad (2.15)$$

Dabei ist I_S ein fiktiver Sperrstrom; für den Korrekturfaktor gilt $a = 1 \dots 2$. Man wählt die Werte für I_S und a empirisch so, dass eine möglichst gute Annäherung der gemessenen Diodenkennlinie an den Funktionsverlauf nach Gl.(2.15) erreicht wird.

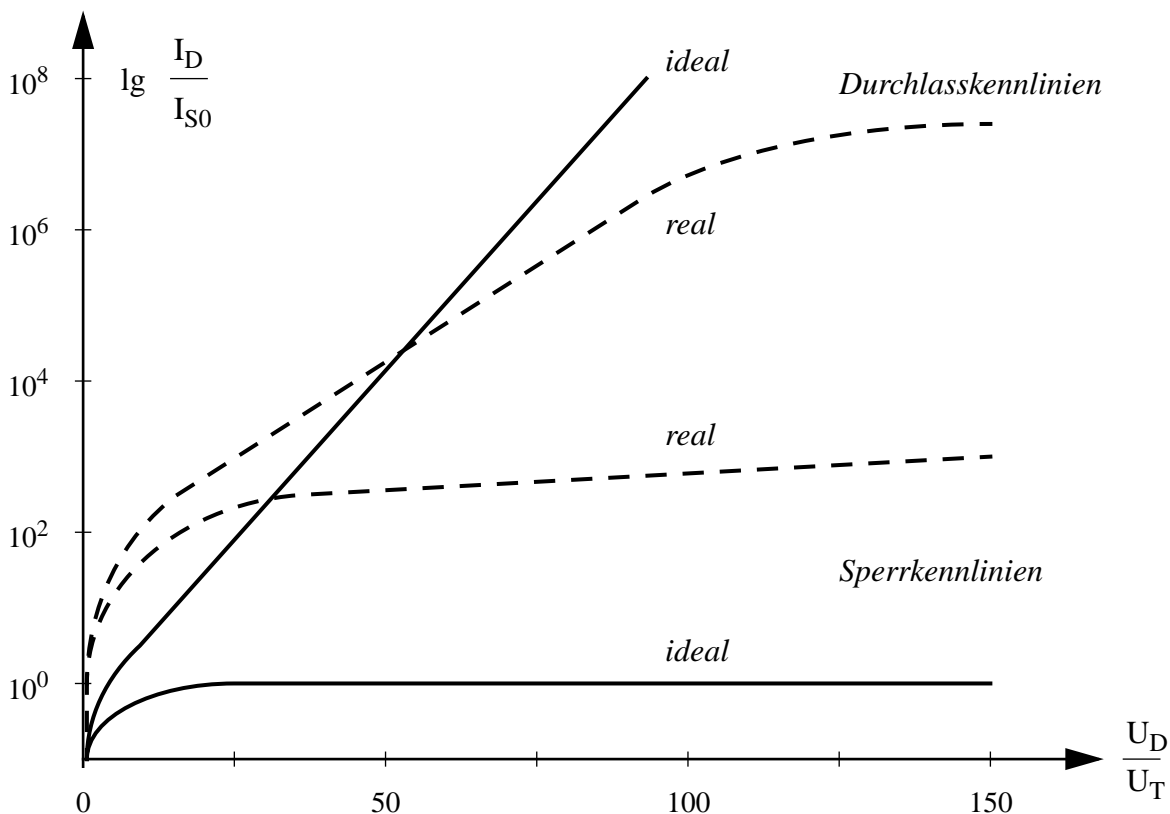


Bild 2.17: Ideale und reale Diodenkennlinien

3.1 Das Überlagerungsprinzip

Hier sollen lineare passive Schaltungen betrachtet werden, d.h. solche, die ausschließlich lineare passive Bauelemente enthalten. Bei linearen Schaltungen gilt das *Überlagerungsprinzip*: Zur Gewinnung einer Lösung wird dem stationären, eingeschwungenen Zustand ein zeitabhängiger, ggf. flüchtiger Ausgleichsvorgang überlagert.

Bild 3.1 stellt technisch wichtige Zeitverläufe einer elektrischen Größe, hier eines Stromes $i(t)$, dar. Man erkennt in den oberen beiden Diagrammen die Überlagerung eines stationären und eines zeitlich veränderlichen Verlaufs. In der digitalen Informationstechnik sind vor allem Impulse mit überlagertem Gleichstrom interessant: Man kann dem Zustand „Stromfluß“ die Binärziffer „1“, dem Zustand „kein Stromfluß“ die Binärziffer „0“ zuordnen (oder auch umgekehrt).

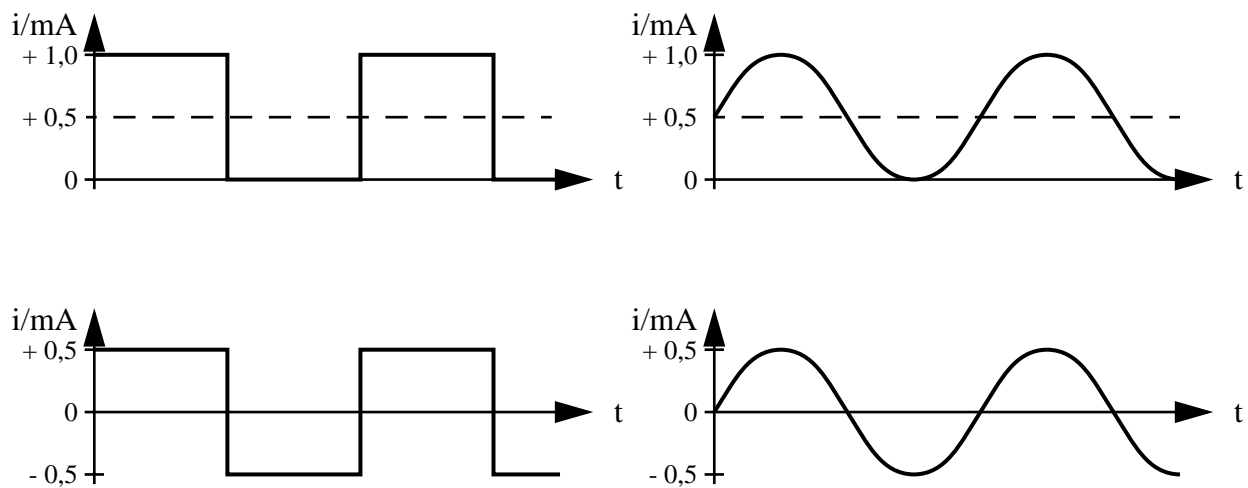


Bild 3.1: Typische Zeitverläufe elektrischer Größen

3.2 Ausgleichsvorgänge im Gleichstromkreis

Die Behandlung von Ausgleichsvorgängen in Gleichstromkreisen geschieht in folgenden Schritten:

- Aufstellen einer Differentialgleichung für den Stromkreis
- Lösen der DGl. für den stationären Zustand
- Lösen der homogenen DGl. für den flüchtigen Vorgang
- Aufstellen der Anfangsbedingung und Überlagerung beider Lösungen

3.2.1 Einschaltvorgang eines Spulenstromes

Nach Bild 3.2 werde eine zunächst stromlose Spule mit der Induktivität L durch Schließen des Schalters S zum Zeitpunkt $t = 0$ über einen Widerstand R an eine Gleichspannung U angeschlossen.

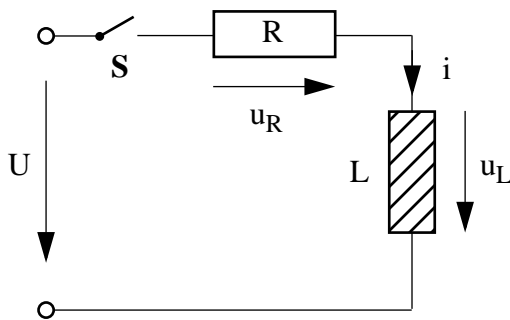


Bild 3.2: Einschalten des Spulenstromes i

Nach der Kirchhoffschen Maschenregel („In jeder Masche ist die Summe der Spannungen gleich Null“) läßt sich in Verbindung mit dem Ohmschen Gesetz Gl.(2.1) und der Spulengleichung Gl.(2.7)

$$u_R = R \cdot i \quad (2.1) \quad u_L = L \cdot \frac{di}{dt} \quad (2.7)$$

für die Schaltung in Bild 3.2 eine „Differentialgleichung“ (DGl.) aufstellen:

$$U = u_R + u_L = R \cdot i + u_L$$

$$U = R \cdot i + L \cdot \frac{di}{dt} \quad \text{Differentialgleichung} \quad (3.1)$$

In einer Schaltung aus linearen Zweipolen erhält man den zeitlichen Verlauf des Spulenstromes $i(t)$ durch die Überlagerung eines stationären („eingeschwungenen“) und eines zeitabhängigen („flüchtigen“) Anteils:

$$i(t) = i_e + i_f(t) \quad (3.2)$$

Im eingeschwungenen Zustand ändert sich der Spulenstrom i nicht mehr:

$$\frac{di}{dt} = 0$$

Damit ergibt sich aus DGl.(3.1) für den eingeschwungenen Anteil:

$$i(t \rightarrow \infty) = i_e = \frac{U}{R} \quad \text{stationäre Lösung} \quad (3.3)$$

während man für den flüchtigen, abklingenden Vorgang aus DGl.(3.1) erhält:

$$0 = R \cdot i_f + L \cdot \frac{di_f}{dt} \quad \text{homogene Differentialgleichung} \quad (3.4)$$

Diese wegen $U = 0$ so genannte „homogene Differentialgleichung“ DGl.(3.4) läßt sich nach Trennung der Variablen integrieren und damit lösen:

$$\int \frac{di_f}{i_f} = -\left(\frac{R}{L}\right) \int dt; \quad \ln i_f = -\left(\frac{R}{L}\right) \cdot t + \text{const.}$$

Man definiert als „Zeitkonstante“ des betrachteten Stromkreises:

$$\tau = \frac{L}{R} \quad (3.5)$$

und erhält so schließlich als flüchtige Lösung der homogenen DGl.(3.4):

$$i_f(t) = \exp\left(\text{const.} - \frac{t}{\tau}\right) = A \cdot \exp\left(-\frac{t}{\tau}\right) \quad \text{flüchtige Lösung} \quad (3.6)$$

Da beim Schließen des Schalters S zur Zeit $t = 0$ der Wert des Stroms durch die zunächst stromlose Spule sich nicht sprunghaft ändern kann (da sich das Magnetfeld in der Spule nicht sprunghaft aufbaut), gilt für $t = 0$ die Anfangsbedingung:

$$i(0) = i_e + i_f(0) = 0 \quad \text{Anfangsbedingung} \quad (3.7)$$

woraus sich für die Konstante A ergibt:

$$0 = \frac{U}{R} + A \quad A = -\left(\frac{U}{R}\right)$$

Man erhält schließlich durch Überlagerung der stationären Lösung Gl.(3.3) mit der flüchtigen Lösung Gl.(3.6) und aufgrund der Anfangsbedingung Gl.(3.7) für den Einschaltvorgang des Spulenstromes:

$$i(t) = \frac{U}{R} \left[1 - \exp\left(-\frac{t}{\tau}\right) \right] \quad \text{Gesamtlösung} \quad (3.8)$$

Mit der nachfolgend nochmals zitierten Spulengleichung Gl.(2.7) erhält man aus dem nach Gl.(3.8) verlaufenden Spulenstrom $i(t)$ den zeitlichen Verlauf der Spulenspannung $u_L(t)$:

$$u_L(t) = L \cdot \frac{di}{dt} \quad (2.7)$$

$$u_L(t) = L \cdot \frac{U}{R} \cdot \left(-\frac{R}{L}\right) \left[-\exp\left(-\frac{t}{\tau}\right)\right]; \quad u_L(t) = U \cdot \exp\left(-\frac{t}{\tau}\right) \quad (3.9)$$

3.2.2 Abklingvorgang eines Spulenstromes

Nach Bild 3.3 werde eine stromdurchflossene Spule mit der Induktivität L durch Schließen des Schalters S zum Zeitpunkt $t = 0$ über einen Widerstand R kurzgeschlossen.

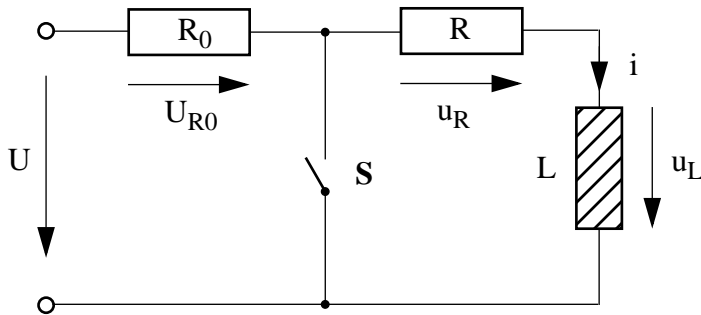


Bild 3.3: Abschalten des Spulenstromes i

Da nach dem Schaltvorgang keine treibende Spannung mehr an der Spule wirkt, gilt dann die homogene DGL.:

$$0 = R \cdot i + L \cdot \frac{di}{dt} \quad \text{homogene Differentialgleichung} \quad (3.10)$$

Den zeitlichen Verlauf des Spulenstroms erhält man wieder durch die Überlagerung eines stationären und eines flüchtigen Anteils:

$$i(t) = i_e + i_f(t)$$

Im stationären Zustand klingt der Spulenstrom völlig ab:

$$i(t \rightarrow \infty) = i_e = 0 \quad \text{stationäre Lösung} \quad (3.11)$$

so daß im folgenden nur der flüchtige Anteil betrachtet werden muß.

Mit der bereits definierten Zeitkonstanten

$$\tau = \frac{L}{R}$$

erhält man nach der im vorhergehenden Abschnitt verwendeten Methode die Lösung der homogenen DGL.(3.10):

$$i(t) = i_f(t) = A \cdot \exp\left(-\frac{t}{\tau}\right) \quad \text{flüchtige Lösung} \quad (3.12)$$

Da beim Schließen des Schalters S zur Zeit $t = 0$ der Wert des Stroms durch die zunächst noch stromdurchflossene Spule sich nicht sprunghaft ändern kann (da sich das Magnetfeld in der Spule nicht sprunghaft abbaut), gilt für $t = 0$ die Anfangsbedingung:

$$i(0) = A = \frac{U}{R_0 + R} \quad \text{Anfangsbedingung} \quad (3.13)$$

Man erhält schließlich durch Einsetzen von Gl.(3.13) in Gl.(3.12) für den Abklingvorgang des Spulenstroms:

$$i(t) = \frac{U}{R_0 + R} \exp\left(-\frac{t}{\tau}\right) \quad \text{Gesamtlösung} \quad (3.14)$$

Mit der bekannten Spulengleichung Gl.(2.7) erhält man aus dem nach Gl.(3.14) verlaufenden Spulenstrom $i(t)$ den zeitlichen Verlauf der Spulenspannung $u_L(t)$:

$$u_L(t) = L \cdot \frac{di}{dt} \quad (2.7)$$

$$u_L(t) = L \cdot \frac{U}{R_0 + R} \cdot \left(-\frac{R}{L}\right) \exp\left(-\frac{t}{\tau}\right); \quad u_L(t) = -U \cdot \frac{R}{R_0 + R} \cdot \exp\left(-\frac{t}{\tau}\right) \quad (3.15)$$

Im Moment des Schaltens $t = 0$ erfolgt an der Spule ein Spannungssprung:

$$u_L(0) = -U \cdot \frac{R}{R_0 + R}$$

3.2.3 Einschaltvorgang einer Kondensatorspannung

Nach Bild 3.4 werde ein zunächst ungeladener Kondensator mit der Kapazität C durch Schließen des Schalters S zum Zeitpunkt $t = 0$ über einen Widerstand R an eine Gleichspannung U angeschlossen.

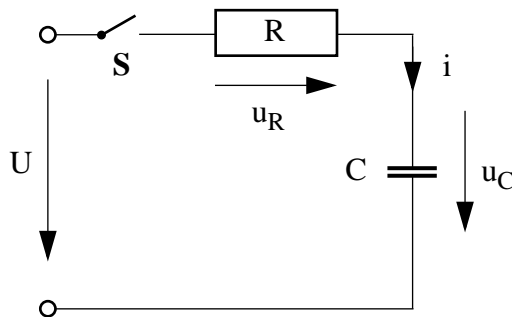


Bild 3.4: Einschalten der Kondensatorspannung u_C

Nach der Kirchhoffschen Maschenregel lässt sich in Verbindung mit dem Ohmschen Gesetz Gl.(2.1) und der Kondensatorgleichung Gl.(2.11):

$$u_R = R \cdot i \quad (2.1) \quad i = C \cdot \frac{du_C}{dt} \quad (2.11)$$

für die Schaltung in Bild 3.4 die folgende Differentialgleichung (DGl.) aufstellen:

$$U = u_R + u_C = R \cdot i + u_C$$

$$U = R \cdot C \cdot \frac{du_C}{dt} + u_C \quad \text{Differentialgleichung} \quad (3.16)$$

In einer Schaltung aus linearen Zweipolen erhält man den zeitlichen Verlauf der Kondensatorspannung $u_C(t)$ durch die Überlagerung eines stationären („eingeschwungenen“) und eines zeitabhängigen („flüchtigen“) Anteils:

$$u_C(t) = u_e + u_f(t) \quad (3.17)$$

Im eingeschwungenen Zustand ändert sich die Kondensatorspannung u_C nicht mehr:

$$\frac{du_C}{dt} = 0$$

Damit ergibt sich aus DGl.(3.16) für den eingeschwungenen Anteil:

$$u_C(t \rightarrow \infty) = u_e = U \quad \text{stationäre Lösung} \quad (3.18)$$

während man für den flüchtigen, abklingenden Vorgang aus DGl.(3.16) erhält:

$$0 = R \cdot C \cdot \frac{du_f}{dt} + u_f \quad \text{homogene Differentialgleichung} \quad (3.19)$$

Die homogene DGl.(3.19) läßt sich nach Trennung der Variablen integrieren und damit lösen:

$$\int \frac{du_f}{u_f} = -\left(\frac{1}{R \cdot C}\right) \int dt; \quad \ln u_f = -\left(\frac{1}{R \cdot C}\right) \cdot t + \text{const.}$$

Man definiert als „Zeitkonstante“ des betrachteten Stromkreises:

$$\tau = R \cdot C \quad (3.20)$$

und erhält so schließlich als flüchtige Lösung der homogenen DGl.(3.19):

$$u_f(t) = \exp\left(\text{const.} - \frac{t}{\tau}\right) = A \cdot \exp\left(-\frac{t}{\tau}\right) \quad \text{flüchtige Lösung} \quad (3.21)$$

Da beim Schließen des Schalters S zur Zeit $t = 0$ der Wert der Spannung am zunächst ungeladenen Kondensator sich nicht sprunghaft ändern kann (da sich das elektrische Feld im Kondensator nicht sprunghaft aufbaut), gilt für $t = 0$ die Anfangsbedingung:

$$u_C(0) = u_e + u_f(0) = 0 \quad \text{Anfangsbedingung} \quad (3.22)$$

woraus sich für die Konstante A ergibt:

$$0 = U + A \quad A = -U$$

Man erhält schließlich durch Überlagerung der stationären Lösung Gl.(3.18) mit der flüchtigen Lösung Gl.(3.21) und aufgrund der Anfangsbedingung Gl.(3.22) für den Einschaltvorgang der Kondensatorspannung:

$$u_C(t) = U \left[1 - \exp\left(-\frac{t}{\tau}\right) \right] \quad \text{Gesamtlösung} \quad (3.23)$$

Mit der nachfolgend zitierten Kondensatorgleichung Gl.(2.11) erhält man aus der nach Gl.(3.23) verlaufenden Kondensatorspannung $u_C(t)$ den zeitlichen Verlauf des Kondensatorstromes $i(t)$:

$$i(t) = C \cdot \frac{du_C}{dt} \quad (2.11)$$

$$i(t) = C \cdot U \cdot \left(-\frac{1}{R \cdot C}\right) \left[-\exp\left(-\frac{t}{\tau}\right)\right]; \quad i(t) = \frac{U}{R} \cdot \exp\left(-\frac{t}{\tau}\right) \quad (3.24)$$

3.2.4 Abklingvorgang einer Kondensatorspannung

Nach Bild 3.5 werde ein geladener Kondensator mit der Kapazität C durch Umlegen des Schalters S zum Zeitpunkt $t = 0$ über einen Widerstand R kurzgeschlossen.

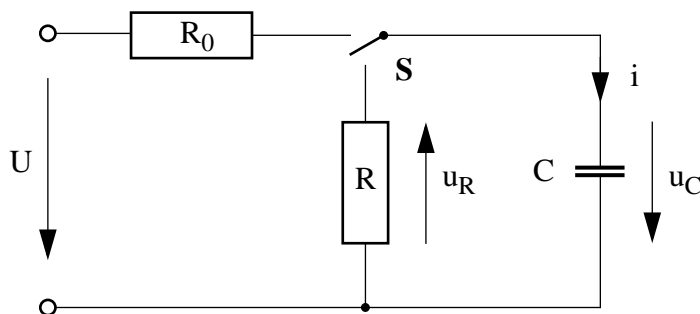


Bild 3.5: Abschalten der Kondensatorspannung u_C

Da nach dem Schaltvorgang keine treibende Spannung mehr am Kondensator wirkt, gilt dann die homogene DGL.:

$$0 = R \cdot C \cdot \frac{du_C}{dt} + u_C \quad \text{homogene Differentialgleichung} \quad (3.25)$$

Den zeitlichen Verlauf der Kondensatorspannung erhält man wieder durch die Überlagerung eines stationären und eines flüchtigen Anteils:

$$u_C(t) = u_e + u_f(t)$$

Im stationären Zustand klingt die Kondensatorspannung völlig ab:

$$u_C(t \rightarrow \infty) = u_e = 0 \quad \text{stationäre Lösung} \quad (3.26)$$

so daß im folgenden nur der flüchtige Anteil betrachtet werden muß.

Mit der bereits definierten Zeitkonstanten

$$\tau = R \cdot C$$

erhält man nach der im vorhergehenden Abschnitt verwendeten Methode die Lösung der homogenen DGL.(3.25):

$$u_C(t) = u_f(t) = A \cdot \exp\left(-\frac{t}{\tau}\right) \quad \text{flüchtige Lösung} \quad (3.27)$$

Da beim Umlegen des Schalters S zur Zeit $t = 0$ der Wert der Spannung am zunächst noch geladenen Kondensator sich nicht sprunghaft ändern kann (da sich das elektrische Feld im Kondensator nicht sprunghaft abbaut), gilt für $t = 0$ die Anfangsbedingung:

$$u_C(0) = A = U \quad \text{Anfangsbedingung} \quad (3.28)$$

Man erhält schließlich durch Einsetzen von Gl.(3.28) in Gl.(3.27) für den Abklingvorgang der Kondensatorspannung:

$$u_C(t) = U \exp\left(-\frac{t}{\tau}\right) \quad \text{Gesamtlösung} \quad (3.29)$$

Mit der bekannten Kondensatorgleichung Gl.(2.11) erhält man aus der nach Gl.(3.29) verlaufenden Kondensatorspannung $u_C(t)$ den zeitlichen Verlauf der Kondensatorstromes $i(t)$:

$$i(t) = C \cdot \frac{du_C}{dt} \quad (2.11)$$

$$i(t) = C \cdot U \cdot \left(\frac{-1}{RC}\right) \exp\left(-\frac{t}{\tau}\right); \quad i(t) = -\left(\frac{U}{R}\right) \cdot \exp\left(-\frac{t}{\tau}\right) \quad (3.30)$$

Im Moment des Umschaltens $t = 0$ erfolgt am Kondensator ein Stromsprung:

$$i(0) = -\left(\frac{U}{R}\right)$$

3.3 Reihenschaltungen im Wechselstrombetrieb

Eine sich zeitlich sinusförmig ändernde physikalische Größe, z.B. ein elektrischer Wechselstrom, der durch seine Amplitude I und seine Kreisfrequenz ω gekennzeichnet ist:

$$i = I \cdot \sin \omega t \qquad \omega = 2\pi f \qquad (3.31)$$

kann formal auch durch einen sich mit der Kreisfrequenz ω im Gegenuhrzeigersinn drehenden „Zeiger“ der Länge I dargestellt werden (Bild 3.6).

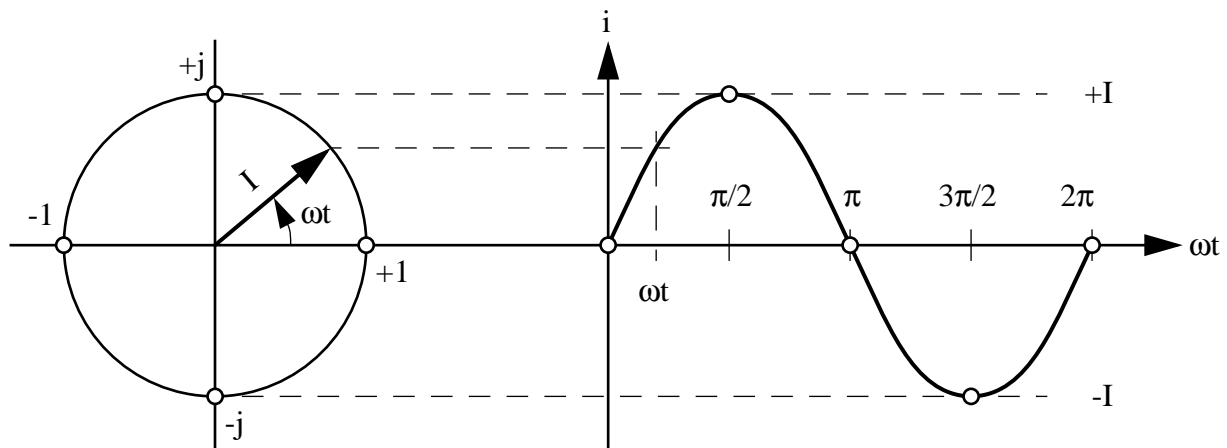


Bild 3.6: Darstellungsformen elektrischer Wechselgrößen

Ferner weiß man, daß nach der Eulerschen Formel gilt:

$e^{j\pi} = -1$	$e^{j2\pi} = +1$	$e^{j\pi/2} = +j$	$e^{-j\pi/2} = 1/j = -j$
links	rechts	oben	unten

Diese „komplexen Größen“ lassen sich in einem rechtwinkligen $[1; j]$ -Koordinatensystem als Zeiger der Länge 1 darstellen, die in die genannten Richtungen weisen. Daraus folgt, daß für einen Wechselstrom die folgende Schreibweise zur obigen Schreibweise nach Gl.(3.31) formal äquivalent ist:

$$\vec{i} = I \cdot e^{j\omega t} \qquad (3.32)$$

Die komplexe Schreibweise nach Gl.(3.32) steht somit neben der Sinusfunktion nach Gl.(3.31) und dem Zeigerdiagramm nach Bild 3.6 als ein weiteres, kompaktes Darstellungsmittel für Wechselstromgrößen zur Verfügung.

3.3.1 Widerstand und Spule in Reihe

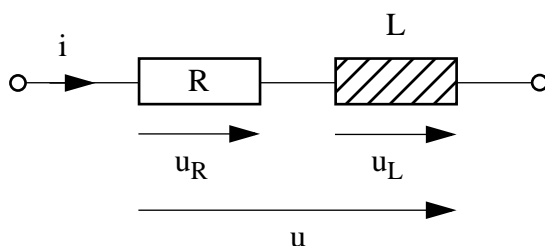


Bild 3.7: Reihenschaltung im Wechselstrombetrieb

Schickt man durch eine Reihenschaltung aus einem Widerstand R und einer Spule L einen Wechselstrom, der sich wie oben erwähnt entweder als Sinusfunktion oder in komplexer Schreibweise darstellen läßt:

$$\begin{aligned} i &= I \cdot \sin \omega t && \text{als Sinusfunktion} \\ \vec{i} &= I \cdot e^{j\omega t} && \text{in komplexer Schreibweise} \end{aligned} \quad (3.33)$$

dann fällt sowohl am Wirkwiderstand $X_R = R$ als auch am Blindwiderstand $X_L = \omega L$ jeweils eine Teilspannung ab. Die einfache Addition dieser Teilspannungen bzw. der Teilwiderstände ist jedoch wegen der an der Spule auftretenden Phasenverschiebung zwischen Strom und Spannung nicht zulässig; vielmehr müssen die genannten Größen als Vektoren mit Betrag und Richtung („Zeiger“) aufgefaßt und vektoriell addiert werden. Anhand von „Zeigerdiagrammen“ soll dies veranschaulicht werden. Die Teilspannung u_R am Widerstand R ist nach dem Ohmschen Gesetz mit dem Wechselstrom i phasengleich:

$$u_R = R \cdot i \quad (2.1)$$

$$\begin{aligned} u_R &= R \cdot I \cdot \sin \omega t = U_R \cdot \sin \omega t && \text{als Sinusfunktion} \\ \vec{u_R} &= R \cdot I \cdot e^{j\omega t} = U_R \cdot e^{j\omega t} && \text{in komplexer Schreibweise} \end{aligned} \quad (3.34)$$

Für die Teilspannung u_L an der Spule L gilt bekanntlich:

$$u_L = L \cdot \frac{di}{dt} \quad (2.7)$$

$$\begin{aligned} u_L &= \omega L \cdot I \cdot \cos \omega t = U_L \cdot \sin(\omega t + \pi/2) && \text{als Sinusfunktion} \\ \vec{u_L} &= j\omega L \cdot I \cdot e^{j\omega t} = U_L \cdot e^{j(\omega t + \pi/2)} && \text{in komplexer Schreibweise} \end{aligned} \quad (3.35)$$

In einer Reihenschaltung fließt durch alle Zweipole derselbe Strom; man verwendet ihn deshalb als Bezugsgröße. Er wird im Zeigerdiagramm als horizontaler Zeiger I dargestellt. Der Wirkspannungszeiger U_R muß deckungsgleich mit dem Stromzeiger I gezeichnet werden, der Blindspannungszeiger U_L dagegen nach Gl.(3.35) um den Phasenwinkel $+\pi/2$ versetzt, d.h senkrecht nach oben, da in der Spule die Spannung bekanntlich dem Strom vorausseilt. Die an der Reihenschaltung abfallende Gesamtspannung U („Scheinspannung“) ergibt sich durch vektorielle Addition der beiden Teilspannungen.

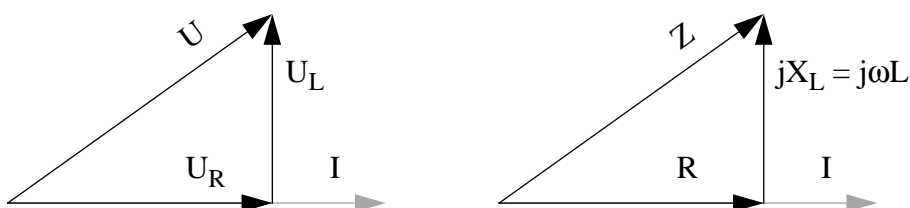


Bild 3.8: Zeigerdiagramm zur Reihenschaltung von Widerstand und Spule

Den Betrag der Scheinspannung U bestimmt man aus dem Zeigerdiagramm wie folgt:

$$U = \sqrt{U_R^2 + U_L^2} = \sqrt{(R \cdot I)^2 + (X_L \cdot I)^2} = I \cdot \sqrt{R^2 + (\omega L)^2}$$

In komplexer Schreibweise:

$$\vec{u} = \vec{u}_R + \vec{u}_L = U_R + jU_L = R \cdot I + jX_L \cdot I = I \cdot (R + j\omega L)$$

Nach dem Ohmschen Gesetz gilt das Zeigerdiagramm nicht nur für die Spannungen, sondern auch für die Widerstände. Damit kann man einen „Scheinwiderstand“ betragsmäßig sowie in komplexer Schreibweise wie folgt definieren:

$$Z = \frac{U}{I} = \sqrt{R^2 + (\omega L)^2} \qquad \vec{Z} = \frac{\vec{u}}{I} = R + j\omega L$$

3.3.2 Widerstand und Kondensator in Reihe

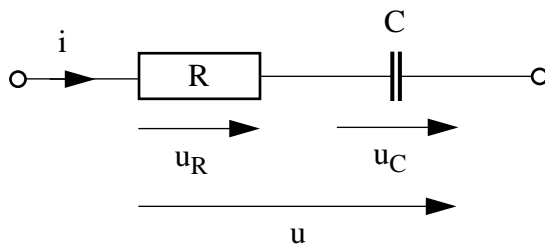


Bild 3.9: Reihenschaltung im Wechselstrombetrieb

Schickt man nun durch eine Reihenschaltung aus einem Widerstand R und einem Kondensator C einen Wechselstrom, der sich bekanntlich entweder als Sinusfunktion oder in komplexer Schreibweise darstellen lässt:

$$\begin{aligned} i &= I \cdot \sin \omega t && \text{als Sinusfunktion} \\ \vec{i} &= I \cdot e^{j\omega t} && \text{in komplexer Schreibweise} \end{aligned} \quad (3.36)$$

dann fällt sowohl am Wirkwiderstand $X_R = R$ als auch am Blindwiderstand $X_C = 1/\omega C$ jeweils eine Teilspannung ab. Auch hier ist eine einfache Addition dieser Teilspannungen bzw. der Teilwiderstände wegen der am Kondensator auftretenden Phasenverschiebung zwischen Strom und Spannung nicht zulässig; die genannten Größen müssen deshalb vektoriell addiert werden. Die Teilspannung u_R am Widerstand R ist nach dem Ohmschen Gesetz mit dem Wechselstrom i phasengleich:

$$u_R = R \cdot i \quad (2.1)$$

$$\begin{aligned} u_R &= R \cdot I \cdot \sin \omega t = U_R \cdot \sin \omega t && \text{als Sinusfunktion} \\ \vec{u}_R &= R \cdot I \cdot e^{j\omega t} = U_R \cdot e^{j\omega t} && \text{in komplexer Schreibweise} \end{aligned} \quad (3.37)$$

Für den Strom i durch den Kondensator C gilt bekanntlich:

$$i = C \cdot \frac{du_C}{dt} \quad (2.11)$$

$$u_C = \frac{-I}{\omega C} \cdot \cos \omega t = U_C \cdot \sin(\omega t - \pi/2) \quad \text{als Sinusfunktion}$$

$$\vec{u}_C = \frac{I}{j\omega C} \cdot e^{j\omega t} = U_C \cdot e^{j(\omega t - \pi/2)} \quad \text{in komplexer Schreibweise} \quad (3.38)$$

Im Zeigerdiagramm muß der Wirkspannungszeiger U_R deckungsgleich mit dem Stromzeiger I , d.h. ebenfalls horizontal eingezeichnet werden, der Blindspannungszeiger U_C dagegen nach Gl.(3.38) um den Phasenwinkel $-\pi/2$ versetzt, d.h. senkrecht nach unten, da im Kondensator die Spannung bekanntlich dem Strom nacheilt. Die an der Reihenschaltung abfallende Gesamtspannung U („Scheinspannung“) ergibt sich wiederum durch vektorielle Addition der Teilspannungen.

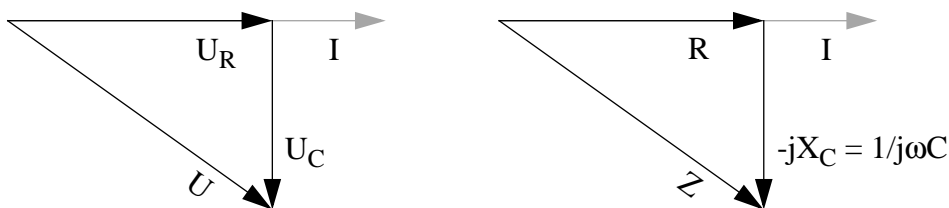


Bild 3.10: Zeigerdiagramm zur Reihenschaltung von Widerstand und Kondensator

Der Betrag der Scheinspannung U errechnet sich aus dem Zeigerdiagramm wie folgt:

$$U = \sqrt{U_R^2 + U_C^2} = \sqrt{(R \cdot I)^2 + (X_C \cdot I)^2} = I \cdot \sqrt{R^2 + (1/\omega C)^2}$$

In komplexer Schreibweise:

$$\vec{u} = \vec{u}_R + \vec{u}_C = U_R - jU_C = R \cdot I - jX_C \cdot I = I \cdot \left(R + \frac{1}{j\omega C} \right)$$

Nach dem Ohmschen Gesetz gilt das Zeigerdiagramm auch für die Widerstände. Damit kann man einen „Scheinwiderstand“ betragsmäßig sowie in komplexer Schreibweise wie folgt definieren:

$$Z = \frac{U}{I} = \sqrt{R^2 + (1/\omega C)^2} \quad \vec{Z} = \frac{\vec{u}}{I} = R + \frac{1}{j\omega C}$$

3.3.3 Der Begriff der Grenzfrequenz

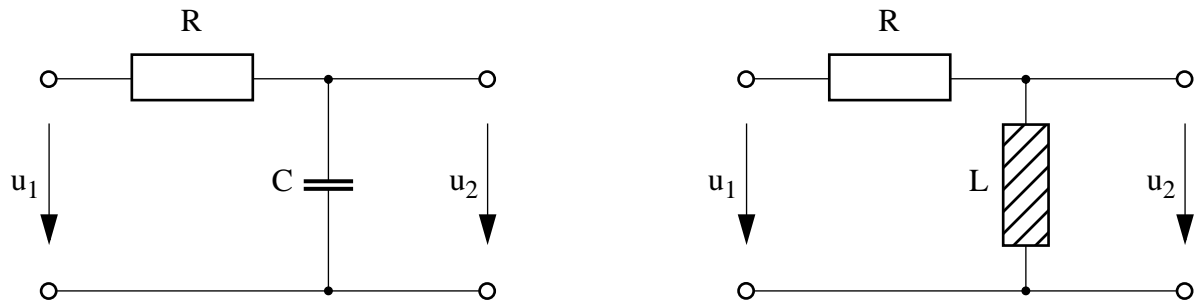


Bild 3.11: Vierpoldarstellung von Spannungsteilern

Man kann eine Reihenschaltung aus Wirk- und Blindwiderstand als frequenzabhängigen Spannungsteiler betrachten und wie in Bild 3.11 als Vierpol darstellen. Wird ein solcher Spannungsteiler mit einer Eingangsspannung konstanter Amplitude aber variabler Frequenz $\omega = 2\pi f$ betrieben, so ist die „Grenzfrequenz“ f_g für den Fall definiert, bei dem die Beträge für Wirk- und Blindwiderstand gleich sind:

$$R = X_C(f_g)$$

$$R = 1 / \omega_g C = 1 / 2\pi f_g C$$

$$R = X_L(f_g)$$

$$R = \omega_g L = 2\pi f_g L$$

Für die Grenzfrequenz gilt dann:

$$f_g = \frac{1}{2\pi \cdot RC}$$

$$f_g = \frac{R}{2\pi \cdot L}$$

(3.39)

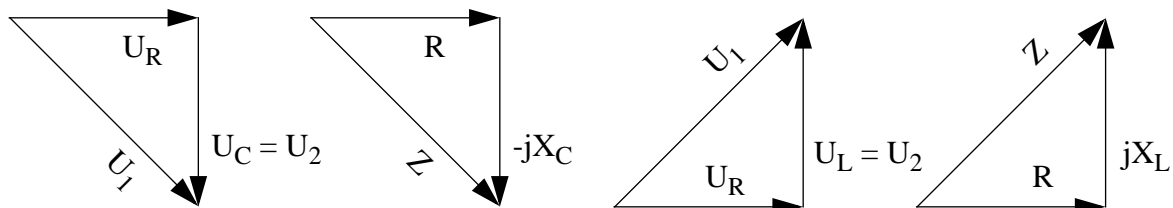


Bild 3.12: Zeigerdiagramme bei der Grenzfrequenz f_g

Eine Vierpolschaltung aus Widerstand R und Kondensator C wird auch als RC-Glied bezeichnet. Für den Frequenzverlauf der Ausgangsspannung $u_2(f)$ des RC-Gliedes erhält man aus dem Widerstandsverhältnis des Spannungsteilers:

$$u_2(f) = u_1 \cdot X_C / \sqrt{R^2 + X_C^2} = u_1 / \sqrt{\left(\frac{R}{X_C}\right)^2 + 1} = u_1 / \sqrt{(2\pi f \cdot RC)^2 + 1}$$

Da nach Gl.(3.39) definitionsgemäß $(2\pi f_g RC) = 1$ gilt, ergibt sich bei der Grenzfrequenz:

$$u_2(f_g) = u_1 / \sqrt{(1)^2 + 1} = u_1 / \sqrt{2} \quad (3.40)$$

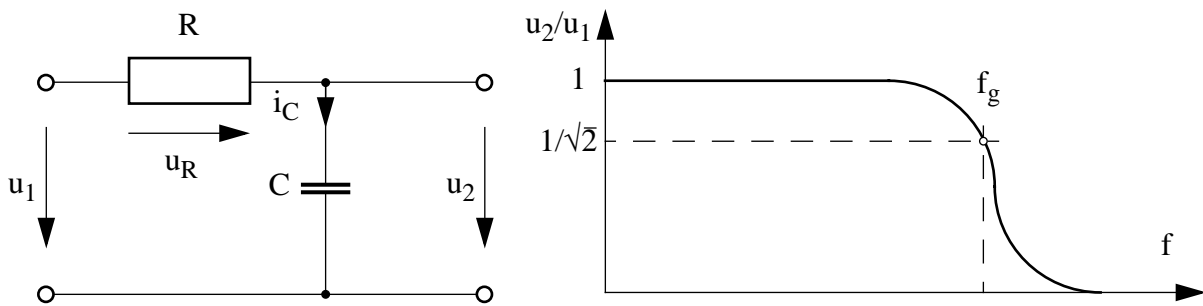


Bild 3.13: Tiefpaß

Das in Bild 3.13 gezeigte RC-Glied wirkt somit beim Betrieb mit sinusförmiger Wechselspannung als „Tiefpaß“, da es tiefe Frequenzen unterhalb seiner Grenzfrequenz durchläßt, aber solche oberhalb derselben stark dämpft, wie der in Bild 3.13 ebenfalls dargestellte Frequenzverlauf zeigt. Dazu komplementär ist der „Hochpaß“, dessen Frequenzverlauf und Schaltung in Bild 3.14 gezeigt wird.

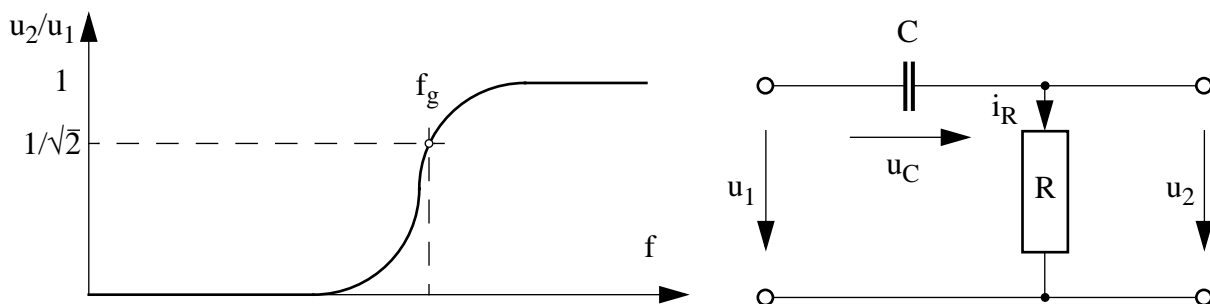


Bild 3.14: Hochpaß

3.3.4 Der Begriff der Resonanzfrequenz

Die „Resonanzfrequenz“ einer Reihen- oder einer Parallelschaltung aus zwei Blindwiderständen, d.h. eines Kondensators C und einer Spule L , ist für den Fall definiert, daß die Beträge der beiden Blindwiderstände gleich sind:

$$X_L(f_0) = X_C(f_0)$$

$$\omega_0 L = 1 / \omega_0 C$$

$$2\pi f_0 L = 1 / 2\pi f_0 C$$

Für die Resonanzfrequenz folgt daraus:

$$f_0 = 1 / 2\pi \sqrt{LC}$$

(3.41)

3.3.5 Der Reihenschwingkreis

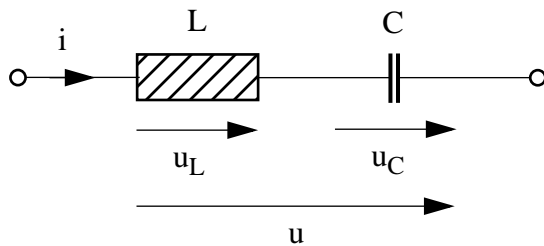


Bild 3.15: Idealer Reihenschwingkreis

Nun werde die Reihenschaltung aus zwei idealen Blindwiderständen $X_L = \omega L$ und $X_C = 1/\omega C$ betrachtet, die von einem gemeinsamen Wechselstrom $i = I \cdot \sin \omega t$ durchflossen werden. Daß in der Spule die Spannung dem Strom voraus-, im Kondensator dagegen nacheilt, wird im Zeigerdiagramm Bild 3.16 durch die Gegenphasigkeit der beiden Blindspannungszeiger U_L und U_C dargestellt. Abhängig von ihren Beträgen ist somit eine teilweise oder auch eine vollständige Kompensation möglich. Dabei sind drei Fälle zu unterscheiden:

- $U_L > U_C$: Es bleibt eine induktive Restspannung $U_L - U_C > 0$ bestehen. Für die verbleibende wirksame Induktivität gilt: $X_{L \text{ rest}} = \omega L - (1/\omega C) = \omega L_{\text{rest}}$
- $U_L = U_C$: In diesem Sonderfall heben sich die beiden Blindspannungen gegenseitig auf, wie man aus dem mittleren Zeigerdiagramm erkennt. Die beiden Blindwiderstände sind betragsgleich („Resonanz“)
- $U_L < U_C$: Es bleibt eine kapazitive Restspannung $U_C - U_L > 0$ bestehen. Für die verbleibende wirksame Kapazität gilt: $X_{C \text{ rest}} = (1/\omega C) - \omega L = 1/\omega C_{\text{rest}}$

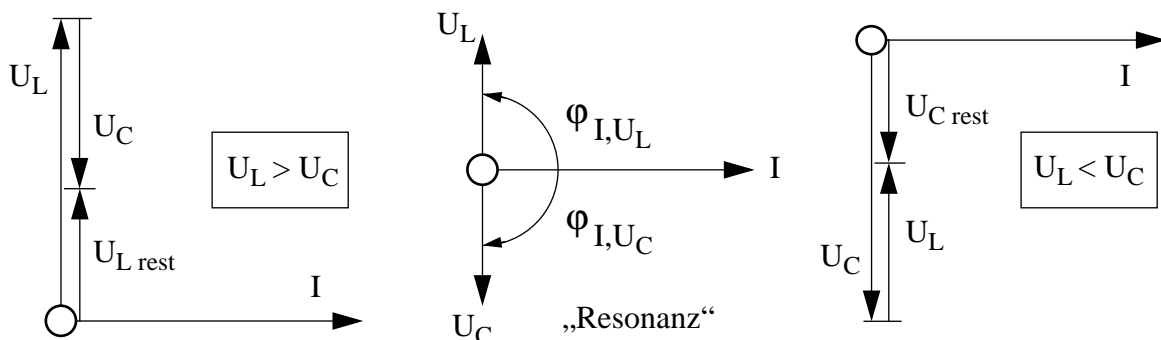


Bild 3.16: Zeigerdiagramme idealer Reihenschwingkreise

Bei Resonanz bedeutet die Kompensation der beiden Blindspannungen im Reihenschwingkreis einen Kurzschluß (!) zwischen seinen beiden Anschlußklemmen: Der Kurzschlußstrom wird theoretisch unendlich groß. Der Wert des Blindwiderstandes bei Resonanz ergibt sich zu:

$$Z_0 = 2\pi f_0 \cdot L = 1/2\pi f_0 \cdot C \quad (3.42)$$

Setzt man in Gl.(3.42) die Resonanzfrequenz f_0 nach Gl.(3.41) ein, so erhält man für den „Kennwiderstand“ des Reihenschwingkreises:

$$Z_0 = \sqrt{L/C} \quad (3.43)$$

In einem aus technischen Bauelementen realisierten Schwingkreis weist die Spule stets einen meßbaren Reihenwiderstand, der Kondensator in der Regel einen nicht zu vernachlässigenden Parallelleitwert auf. Diese sog. „parasitären“, weil störenden Größen können in einem in Reihe liegenden Ersatzwiderstand zusammengefaßt werden wie Bild 3.17 zeigt: Ein realer, verlustbehafteter Reihenschwingkreis besteht aus einer Reihenschaltung von Widerstand R, Spule L und Kondensator C.

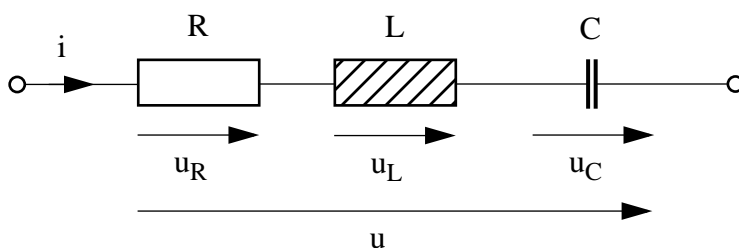


Bild 3.17: Verlustbehafteter Reihenschwingkreis

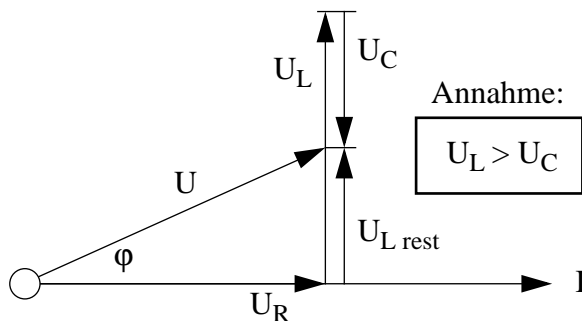


Bild 3.18: Zeigerdiagramm des verlustbehafteten Reihenschwingkreises

3.4 Parallelschaltungen im Wechselstrombetrieb

3.4.1 Widerstand und Kondensator parallel

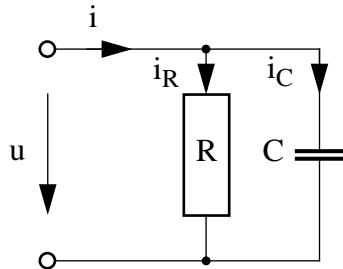


Bild 3.19: Parallelschaltung im Wechselstrombetrieb

Legt man an eine Parallelschaltung aus einem Leitwert $G = 1/R$ und einem Kondensator C eine Wechselspannung

$$\begin{aligned} u &= U \cdot \sin \omega t && \text{als Sinusfunktion} \\ \vec{u} &= U \cdot e^{j\omega t} && \text{in komplexer Schreibweise} \end{aligned} \quad (3.44)$$

so fließt sowohl durch den Wirkleitwert $Y_R = G$ als auch durch den Blindleitwert $Y_C = \omega C$ jeweils ein Teilstrom. Die einfache Addition dieser Teilströme bzw. der Teilleitwerte ist jedoch wegen der am Kondensator auftretenden Phasenverschiebung zwischen Strom und Spannung nicht zulässig; vielmehr müssen die genannten Größen als Vektoren mit Betrag und Richtung („Zeiger“) aufgefaßt und vektoriell addiert werden. Bei der Berechnung einer Schaltung im Wechselstrombetrieb ist daher als erstes ein Zeigerdiagramm zu erstellen. Der Teilstrom i_R durch den Leitwert $G = 1/R$ ist nach dem Ohmschen Gesetz mit der angelegten Wechselspannung u phasengleich:

$$\begin{aligned} (2.1) \quad i_R &= G \cdot u \\ i_R &= G \cdot U \cdot \sin \omega t = I_R \cdot \sin \omega t && \text{als Sinusfunktion} \\ \vec{i_R} &= G \cdot U \cdot e^{j\omega t} = I_R \cdot e^{j\omega t} && \text{in komplexer Schreibweise} \end{aligned} \quad (3.45)$$

Für den Teilstrom i_C durch den Kondensator C gilt bekanntlich:

$$\begin{aligned} i_C &= C \cdot \frac{du}{dt} && (2.11) \\ i_C &= \omega C \cdot U \cdot \cos \omega t = I_C \cdot \sin(\omega t + \pi/2) && \text{als Sinusfunktion} \\ \vec{i_C} &= j\omega C \cdot U \cdot e^{j\omega t} = I_C \cdot e^{j(\omega t + \pi/2)} && \text{in komplexer Schreibweise} \end{aligned} \quad (3.46)$$

In einer Parallelschaltung liegt an allen Zweipolen dieselbe Spannung; man verwendet sie deshalb als Bezugsgröße. Sie wird im Zeigerdiagramm als horizontaler Zeiger U dargestellt. Der Wirkstromzeiger I_R muß deckungsgleich mit dem Spannungszeiger U gezeichnet werden, der Blindstromzeiger I_C dagegen nach Gl.(3.46) um den Phasenwinkel $+\pi/2$ versetzt, d.h. senkrecht nach oben, da

im Kondensator der Strom bekanntlich der Spannung vorausleitet. Der durch die Parallelschaltung fließende Gesamtstrom I („Scheinstrom“) ergibt sich durch vektorielle Addition der Teilströme.

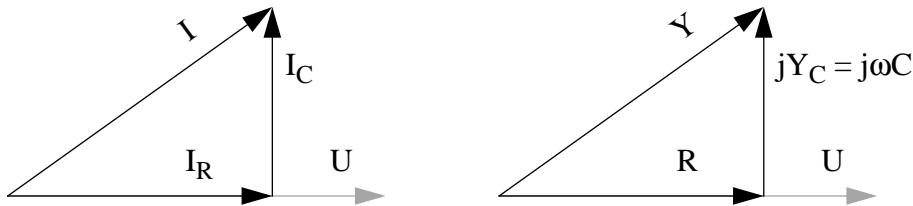


Bild 3.20: Zeigerdiagramm zur Parallelschaltung von Widerstand und Kondensator

Der Betrag des Scheinstroms I bestimmt man aus dem Zeigerdiagramm wie folgt:

$$I = \sqrt{I_R^2 + I_C^2} = \sqrt{(G \cdot U)^2 + (Y_C \cdot U)^2} = U \cdot \sqrt{G^2 + (\omega C)^2}$$

In komplexer Schreibweise:

$$\vec{i} = \vec{i}_R + \vec{i}_C = I_R + jI_C = G \cdot U + jY_C \cdot U = U \cdot (G + j\omega C)$$

Nach dem Ohmschen Gesetz gilt das Zeigerdiagramm nicht nur für die Ströme, sondern auch für die Leitwerte. Damit kann man den „Scheinleitwert“ betragsmäßig sowie in komplexer Schreibweise wie folgt definieren:

$$Y = \frac{I}{U} = \sqrt{G^2 + (\omega C)^2} \qquad \vec{Y} = \frac{\vec{i}}{U} = G + j\omega C$$

3.4.2 Widerstand und Spule parallel

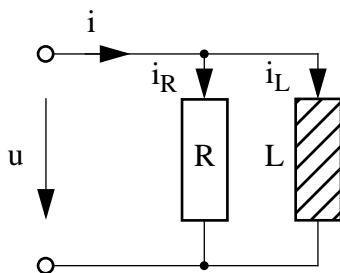


Bild 3.21: Parallelschaltung im Wechselstrombetrieb

Legt man an eine Parallelschaltung aus einem Leitwert $G = 1/R$ und einer Spule L eine Wechselspannung, der sich bekanntlich entweder als Sinusfunktion oder in komplexer Schreibweise darstellen lässt:

$$\begin{aligned} u &= U \cdot \sin \omega t && \text{als Sinusfunktion} \\ \vec{u} &= U \cdot e^{j\omega t} && \text{in komplexer Schreibweise} \end{aligned} \quad (3.47)$$

dann fließt sowohl durch den Wirkleitwert $Y_R = G$ und den Blindleitwert $Y_L = 1/\omega L$ jeweils ein Teilstrom. Auch hier ist eine einfache Addition dieser Teilströme bzw. der Teilleitwerte wegen der

an der Spule auftretenden Phasenverschiebung zwischen Strom und Spannung nicht zulässig; die genannten Größen müssen deshalb vektoriell addiert werden. Der Teilstrom i_R durch den Leitwert $G = 1/R$ ist nach dem Ohmschen Gesetz mit der Wechselspannung u phasengleich:

$$i_R = G \cdot u \quad (2.1)$$

$$i_R = G \cdot U \cdot \sin \omega t = I_R \cdot \sin \omega t \quad \text{als Sinusfunktion}$$

$$\vec{i}_R = G \cdot U \cdot e^{j\omega t} = I_R \cdot e^{j\omega t} \quad \text{in komplexer Schreibweise} \quad (3.48)$$

Für die Spannung u an der Spule L gilt bekanntlich:

$$u = L \cdot \frac{di_L}{dt} \quad (2.7)$$

$$i_L = \frac{-U}{\omega L} \cdot \cos \omega t = I_L \cdot \sin(\omega t - \pi/2) \quad \text{als Sinusfunktion}$$

$$\vec{i}_L = \frac{U}{j\omega L} \cdot e^{j\omega t} = I_C \cdot e^{j(\omega t - \pi/2)} \quad \text{in komplexer Schreibweise} \quad (3.49)$$

Im Zeigerdiagramm muß der Wirkstromzeiger I_R deckungsgleich mit dem Spannungszeiger U , d.h. ebenfalls horizontal gezeichnet werden, der Blindstromzeiger I_L dagegen nach Gl.(3.49) um den Phasenwinkel $-\pi/2$ versetzt, d.h. senkrecht nach unten, da in der Spule der Strom bekanntlich der Spannung nacheilt. Der durch die Parallelschaltung fließende Gesamtstrom I („Scheinstrom“) ergibt sich wiederum durch vektorielle Addition der Teilströme.

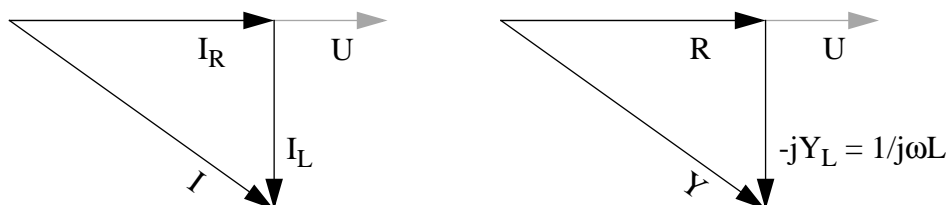


Bild 3.22: Zeigerdiagramm zur Parallelschaltung von Widerstand und Spule

Der Betrag des Scheinstroms I errechnet sich aus dem Zeigerdiagramm wie folgt:

$$I = \sqrt{I_R^2 + I_L^2} = \sqrt{(G \cdot U)^2 + (Y_L \cdot U)^2} = U \cdot \sqrt{(1/R)^2 + (1/\omega L)^2}$$

In komplexer Schreibweise:

$$\vec{i} = \vec{i}_R + \vec{i}_L = I_R - jI_L = G \cdot U - jY_L \cdot U = U \cdot \left(\frac{1}{R} + \frac{1}{j\omega L} \right)$$

Das Zeigerdiagramm gilt, wie oben erwähnt, auch für die Leitwerte. Damit kann man einen „Scheinleitwert“ betragsmäßig sowie in komplexer Schreibweise wie folgt definieren:

$$Y = \frac{I}{U} = \sqrt{G^2 + (1/\omega L)^2} \quad \vec{Y} = \frac{\vec{i}}{U} = \frac{1}{R} + \frac{1}{j\omega L}$$

3.4.3 Der Parallelschwingkreis

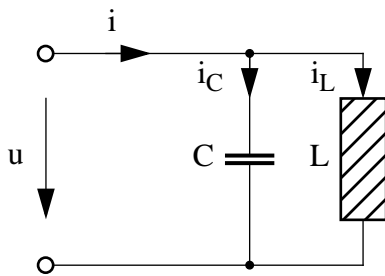


Bild 3.23: Idealer Parallelschwingkreis

Nun werde die Parallelschaltung aus zwei idealen Blindleitwerten $Y_C = \omega C$ und $Y_L = 1/\omega L$ betrachtet, an der eine gemeinsame Wechselspannung $u = U \cdot \sin \omega t$ anliegt. Daß im Kondensator der Strom der Spannung voraus-, in der Spule dagegen nacheilt, wird im Zeigerdiagramm Bild 3.24 durch die Gegenphasigkeit der beiden Blindstromzeiger I_C und I_L dargestellt. Abhängig von ihren Beträgen ist somit eine teilweise oder auch eine vollständige Kompensation möglich. Dabei sind drei Fälle zu unterscheiden:

- $I_C > I_L$: Es bleibt ein kapazitiver Reststrom $I_C - I_L > 0$ bestehen. Für die verbleibende wirksame Kapazität gilt: $Y_{C \text{ rest}} = \omega C - (1/\omega L) = \omega C_{\text{rest}}$
- $I_C = I_L$: In diesem Sonderfall heben sich die beiden Blindströme gegenseitig auf, wie man aus dem mittleren Zeigerdiagramm erkennt. Die beiden Blindleitwerte sind betragsgleich („Resonanz“)
- $I_C < I_L$: Es bleibt ein induktiver Reststrom $I_L - I_C > 0$ bestehen. Für die verbleibende wirksame Induktivität gilt: $Y_{L \text{ rest}} = (1/\omega L) - \omega C = 1/\omega L_{\text{rest}}$

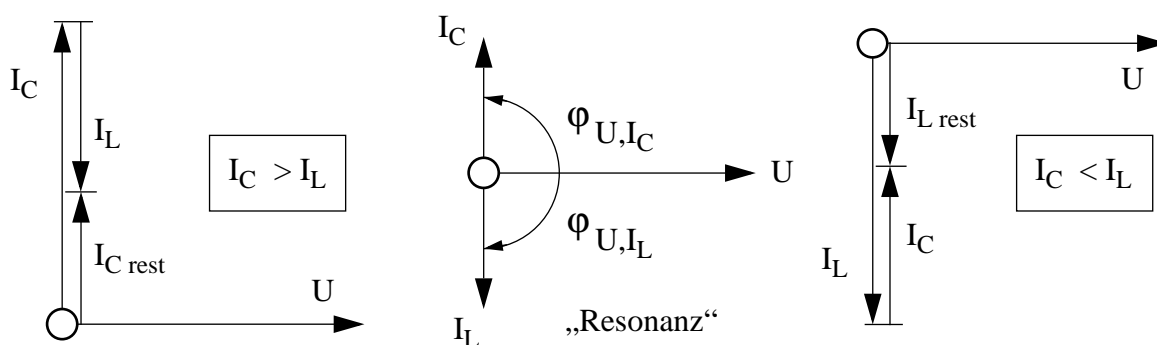


Bild 3.24: Zeigerdiagramme idealer Parallelschwingkreise

Bei Resonanz bedeutet die Kompensation der beiden Blindströme im Parallelschwingkreis einen Leerlauf (!) zwischen seinen beiden Anschlußklemmen: Die Leerlaufspannung wird theoretisch unendlich groß. Der Wert des Blindleitwerts bei Resonanz ergibt sich zu:

$$Y_0 = 2\pi f_0 \cdot C = 1/2\pi f_0 \cdot L \quad (3.50)$$

Setzt man in Gl.(3.50) die Resonanzfrequenz f_0 nach Gl.(3.41) ein, so erhält man für den „Kennwiderstand“ des Parallelschwingkreises dasselbe Ergebnis wie im Fall des Reihenschwingkreises:

$$Z_0 = 1/Y_0 = \sqrt{L/C} \quad (3.43)$$

Wie erwähnt weist in einem aus technischen Bauelementen realisierten Schwingkreis die Spule stets einen meßbaren Reihenwiderstand, der Kondensator in der Regel einen nicht zu vernachlässigenden Parallelleitwert auf. Diese „parasitären“ Größen können in einem parallel liegenden Ersatzwiderstand zusammengefaßt werden wie Bild 3.25 zeigt: Ein realer, verlustbehafteter Parallelschwingkreis besteht aus einer Parallelschaltung von Widerstand R , Spule L und Kondensator C .

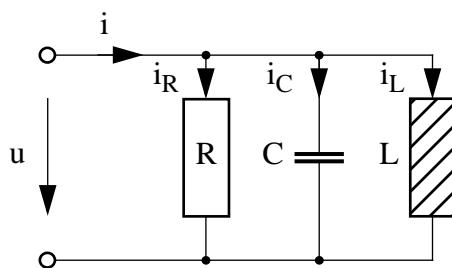


Bild 3.25: Verlustbehafteter Parallelschwingkreis

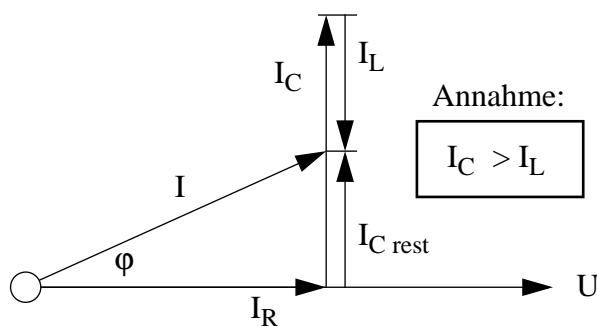


Bild 3.26: Zeigerdiagramm des verlustbehafteten Parallelschwingkreises

3.5 RC-Glieder im Impulsbetrieb

3.5.1 Das Differenzierglied

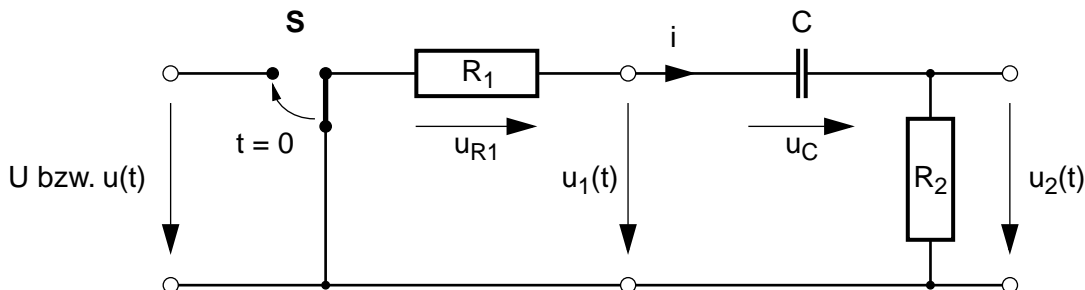


Bild 3.27: Differenzierglied aus linearen Zweipolen

In der im obigen Bild gezeigten Schaltung werde zum Zeitpunkt $t = 0$ der Schalter S umgeschaltet, d.h. die Gleichspannungsquelle U wird an die Schaltung angeschlossen. Nach der Kirchhoffschen Maschenregel gilt dann unter Verwendung der bereits bekannten Kondensatorgleichung Gl.(2.11):

$$i = C \cdot \frac{du_C}{dt} \quad (2.11)$$

$$U = u_{R1} + u_C + u_2 = R_1 \cdot i + \frac{1}{C} \int i dt + R_2 \cdot i \quad (3.51)$$

Differenziert man Gl.(3.51) nach t , so erhält man die homogene Differentialgleichung (DGL):

$$0 = R_1 \cdot \frac{di}{dt} + \frac{i}{C} + R_2 \cdot \frac{di}{dt} \quad \text{Differentialgleichung} \quad (3.52)$$

Durch Trennung der Variablen wird die Integration und damit die Lösung der DGL.(3.52) möglich:

$$\int \frac{di}{i} = - \int \frac{dt}{(R_1 + R_2)C} \quad \ln i = - \frac{t}{(R_1 + R_2)C} + \text{const.} \quad (3.53)$$

Man definiert als „Zeitkonstante“ dieses RC-Glieds:

$$\tau = (R_1 + R_2)C \quad (3.54)$$

und erhält aus Gl.(3.53) mit dieser Zeitkonstanten als Lösung der DGL.(3.52):

$$i(t) = \exp \left[- \left(\frac{t}{\tau} \right) + \text{const.} \right] = A \cdot e^{-t/\tau} \quad (3.55)$$

Da beim Schließen des Schalters S zur Zeit $t = 0$ die Spannung $u(0) = 0$ am zunächst ungeladenen Kondensator C sich nicht sprunghaft ändern kann, liest man aus dem Schaltbild als Anfangsbedingung ab:

$$i(0) = \frac{U}{R_1 + R_2} = A \quad \text{Anfangsbedingung} \quad (3.56)$$

So erhält man schließlich für den zeitlichen Verlauf des Stroms $i(t)$ durch das RC-Glied:

$$i(t) = \frac{U}{R_1 + R_2} \cdot e^{-t/\tau} \quad \text{Lösung der DGL.(3.52)} \quad (3.57)$$

Der Strom $i(t)$ nach Gl.(3.57) bestimmt die Ausgangsspannung des RC-Gliedes, die am Widerstand R_2 abfällt:

$$u_2(t) = R_2 \cdot i(t) = \frac{R_2 \cdot U}{R_1 + R_2} \cdot e^{-t/\tau} \quad (3.58)$$

Nun werde an die Eingangsklemmen des RC-Glieds eine Spannung $u(t)$ angelegt, die sich in Form von Rechteckimpulsen zwischen dem Wert Null und dem Wert der Gleichspannung U periodisch ändert (wobei der Schalter S stets geschlossen bleibt). Wählt man die Zeitkonstante des RC-Glieds klein gegenüber der Periodendauer dieser Spannungsimpulse ($\tau \ll T, T_1, T_2$), so entstehen an seinen Ausgangsklemmen „Nadelimpulse“ $u_2(t)$ wie in Bild 3.28 dargestellt. Dieses Verhalten soll im folgenden hergeleitet werden.

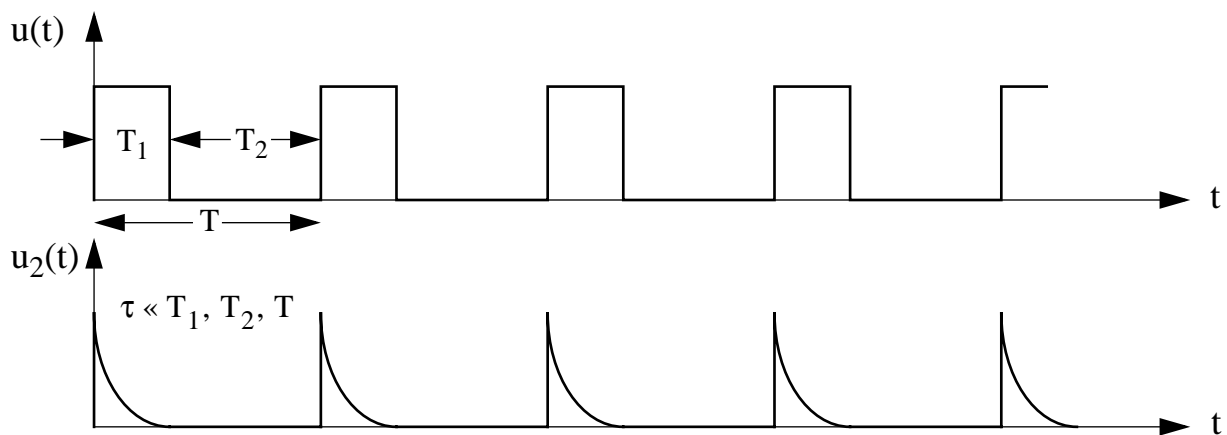


Bild 3.28: Ein- und Ausgangsspannung des Differenzierglieds

Nimmt man an, die Eingangsspannung $u(t)$ habe eine obere Grenzfrequenz ω_g und dimensioniert man das RC-Glied wie folgt:

$$\tau \ll \frac{1}{\omega_g} \quad R_1 + R_2 \ll \frac{1}{\omega_g C} \quad (3.59)$$

dann ist für jede auftretende Frequenz $\omega \leq \omega_g$ der Blindwiderstand $1/\omega C$ des Kondensators sehr viel größer als die Summe der Wirkwiderstände ($R_1 + R_2$). Die angelegte Spannung fällt dann praktisch nur am Kondensator C ab, so daß der Strom fast ausschließlich von ihm bestimmt wird. Dann gilt näherungsweise:

$$i(t) = C \frac{du_C}{dt} \approx C \frac{du_1}{dt} \approx C \frac{d}{dt} u(t) \quad (3.60)$$

woraus für die Ausgangsspannung folgt:

$$u_2(t) \approx R_2 C \cdot \frac{d}{dt} u(t) \quad \text{da } u_2(t) = R_2 \cdot i(t) \quad (3.61)$$

Man erkennt, daß das in Bild 3.27 gezeigte RC-Glied in 1. Näherung eine Differentiation der Eingangsimpulse $u(t)$ bewirkt; es wird deshalb als „Differenzierglied“ bezeichnet. (Wird diesselbe Schaltung mit sinusförmiger Wechselspannung betrieben, so wirkt sie als Hochpaß.)

3.5.2 Das Integrierglied

Die im folgenden Bild gezeigte Schaltung unterscheidet sich vom Differenzierglied, das im vorhergehenden Abschnitt vorgestellt wurde, durch die Vertauschung von Widerstand R_2 und Kondensator C . Zum Zeitpunkt $t = 0$ werde der Schalter S umgeschaltet, d.h. die Gleichspannungsquelle U wird an die Schaltung angeschlossen.

Die für das Differenzierglied durchgeführte Berechnung des Gesamtstromes $i(t)$ durch das RC-Glied ist unabhängig von der Reihenfolge der Zweipole und deshalb auch hier gültig:

$$(3.57) \quad i(t) = \frac{U}{R_1 + R_2} \cdot e^{-t/\tau} \quad \text{Lösung der DGL.(3.52)}$$

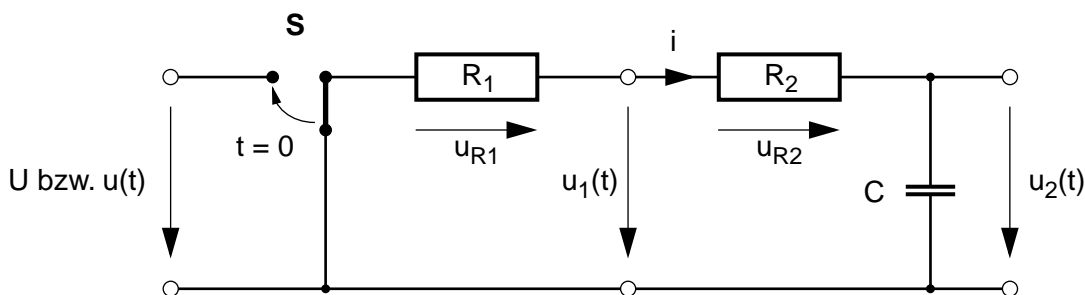


Bild 3.29: Integrierglied aus linearen Zweipolen

Nach der Kirchhoffschen Maschenregel gilt für die Ausgangsspannung des in Bild 3.29 gezeigten RC-Glieds:

$$u_2(t) = U - (R_1 + R_2) \cdot i(t) \quad (3.62)$$

Einsetzen des Stromes $i(t)$ nach Gl.(3.57) ergibt:

$$u_2(t) = U \cdot (1 - e^{-t/\tau}) \quad (3.63)$$

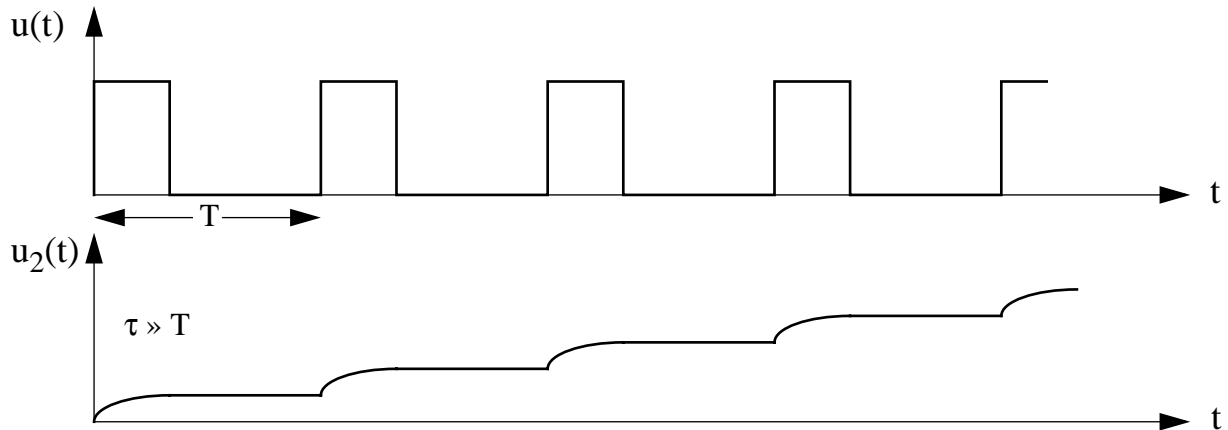


Bild 3.30: Ein- und Ausgangsspannung des Integrierglieds

Nun werde an die Eingangsklemmen des RC-Glieds eine Spannung $u(t)$ angelegt, die sich in Form von Rechteckimpulsen zwischen dem Wert Null und dem Wert der Gleichspannung U periodisch ändert (wobei der Schalter S stets geschlossen bleibt). Wählt man die Zeitkonstante des RC-Glieds groß gegenüber der Periodendauer dieser Spannungsimpulse ($\tau \gg T$), so entsteht an seinem Ausgang eine monoton ansteigende „Treppenspannung“ $u_2(t)$ wie in Bild 3.30 dargestellt. Dieses Verhalten soll im folgenden hergeleitet werden.

Nimmt man an, die Eingangsspannung $u(t)$ habe eine untere Grenzfrequenz ω_u und dimensioniert man das RC-Glied wie folgt:

$$\tau \gg \frac{1}{\omega_u} \qquad R_1 + R_2 \gg \frac{1}{\omega_u C} \qquad (3.64)$$

So ist für jede auftretende Frequenz $\omega \geq \omega_u$ die Summe der Wirkwiderstände ($R_1 + R_2$) sehr viel größer als der Blindwiderstand $1/\omega C$ des Kondensators. Die angelegte Spannung fällt dann praktisch nur an den beiden Widerständen ab, so daß der Strom fast ausschließlich von ihnen bestimmt wird. Dann gilt näherungsweise:

$$i(t) \approx \frac{u(t)}{R_1 + R_2} \qquad (3.65)$$

woraus durch Umstellen der bekannten Kondensatorgleichung für die Ausgangsspannung folgt:

$$u_2(t) = \frac{1}{C} \int i(t) dt = \frac{1}{\tau} \int u(t) dt \qquad \text{da } i = C \cdot \frac{du_2}{dt} \qquad (3.66)$$

Man erkennt, daß das in Bild 3.29 gezeigte RC-Glied in 1. Näherung eine Integration der Eingangsimpulse $u(t)$ bewirkt; es wird deshalb als „Integrierglied“ bezeichnet. (Wird diesselbe Schaltung mit sinusförmiger Wechselspannung betrieben, so wirkt sie als Tiefpaß.)

Der elektronische Teil digitaler Rechenanlagen wird mit höchstintegrierten Schaltungen aufgebaut. Deren Herstellung beruht auf theoretischen Erkenntnissen der Festkörperphysik und den praktischen Erfahrungen der Halbleitertechnologie. Daher gibt dieses Kapitel einen Überblick über die Grundbegriffe der Stromleitung in Halbleitern. Wir werden uns dazu mit den elektrischen Effekten in Halbleiterkristallen befassen und mit den Modellvorstellungen, die zu ihrer Deutung entwickelt wurden. Im Gegensatz zur Stromleitung in Metallen wird die Stromleitung in Halbleitern nicht vom Halbleiterkristall selbst, sondern von den *Störungen* des Kristalls bestimmt: Es ist vom idealen, ungestörten Kristallgitter auszugehen, in das gezielt Störstellen eingebracht werden. Da ein Kristall durch die Verbindung von Atomen entsteht, werden wir uns mit den Kristallatomen und ihren Elektronen befassen, und zwar am Beispiel des fast ausschließlich verwendeten Halbleitermaterials Silizium.

4.1 Grundlagen des Halbleitereffekts

Zur Veranschaulichung elektronischer Leitungsvorgänge in Halbleitern dienen zwei Modellvorstellungen, die sich ergänzen:

- Das „Kristallgitter“. Das Gittermodell geht vom Elektron als Korpuskel aus und stellt die geometrische Anordnung der Atome und ihrer Elektronen in einem Halbleiterkristall schematisch dar.
- Die „Energiebänder“. Das Bändermodell kann aus der Wellennatur des Elektrons hergeleitet werden; es sagt etwas über die Energie der Elektronen in einem Halbleiterkristall aus.

Die chemischen Elemente, deren elektrische Eigenschaften man mittels Gitter- und Bändermodell erklären kann, falls sie in kristalliner Form vorliegen, findet man in der IV. Gruppe des Periodischen Systems der Elemente (Bild 4.1).

3 Valenzelektronen III a		4 Valenzelektronen IV a		5 Valenzelektronen V a	
	Bor B 5		Kohlenstoff C 6		Stickstoff N 7
	Aluminium Al 13		Silizium Si 14		Phosphor P 15
Skandium Sc 21		Titan Ti 22		Vanadium V 23	
	Gallium Ga 31		Germanium Ge 32		Arsen As 33
Yttrium Y 39		Zirkonium Zr 40		Niob Nb 41	
	Indium In 49		Zinn Sn 50		Antimon Sb 51

Bild 4.1: Ausschnitt aus dem Periodischen System der Elemente.
Die Halbleiter findet man in der Gruppe IV b.

- *Diamant*, die kristalline Form des Kohlenstoffs, und *Zinn*, das als graues Zinn ebenfalls in einem Gitter vom Diamanttyp kristallisiert, scheiden zur Herstellung elektronischer Schaltungen von vornherein aus: der Diamant, weil er sich selbst bei stärkster Energiezufuhr nach wie vor wie ein Isolator verhält; das graue Zinn, weil es bei Zimmertemperatur bereits so stark wie ein Metall leitet.
- Von den beiden dazwischen liegenden echten Halbleitern hat sich das *Silizium* gegenüber dem *Germanium* wegen eines technologischen Vorteils weitgehend durchgesetzt. Es kann durch einen Oxidationsvorgang relativ einfach mit einer sehr widerstandsfähigen Quarzglasschicht (SiO_2) überzogen werden, was erst die Anwendung photolithographischer Methoden bei der Herstellung integrierter Schaltkreise ermöglicht hat.

4.1.1 Das Kristallgittermodell

Das Gittermodell schematisiert die geometrische Anordnung der Atome und Elektronen, die sich zu einem Kristall verbunden haben, in einer ebenen Darstellung. Die Bindungskräfte, die einen Halbleiterkristall zusammenhalten, gleichen denen, die auftreten, wenn sich zum Beispiel zwei Wasserstoffatome 2H zu einem Wasserstoffmolekül H_2 verbinden (Bild 4.2). Beide Atome bringen je ein Elektron mit, die sich zu einem Elektronenpaar vereinigen, beide Atomkerne gemeinsam umkreisen und so zu einer chemischen „Valenzbindung“ führen.



Bild 4.2: Valenzbindung des Wasserstoffmoleküls H_2
links: Atommodell; rechts: schematische Darstellung

Die Elementhalbleiter sind chemisch 4-wertig, das heißt, ein Siliziumatom besitzt in der äußersten Elektronenschale vier Valenzelektronen. Da die Kristallbindung nur durch sie zustande kommt, können die übrigen Elektronen zusammen mit dem Atomkern zu einem 4-fach positiven „Atomrumpf“ zusammengefaßt werden, der von den vier Valenzelektronen umkreist wird (Bild 4.3 links). Verbinden sich die Siliziumatome zu einem Kristall, so tauscht jedes (4-wertige) Kristallatom mit vier Nachbaratomen je ein Valenzelektron aus, die sich zu Elektronenpaaren vereinigen und jeweils zwei direkt benachbarte Atome gemeinsam umkreisen (Bild 4.3 rechts). So kann jedes Siliziumatom durch das Zusammenwirken mit vier Nachbaratomen seine vier chemischen Valenzen absättigen und damit einen stabilen Zustand einnehmen.

- Sind alle Valenzelektronen im Kristallgitter gebunden, so steht keines für eine Stromleitung zur Verfügung: Der ungestörte, ideale Halbleiterkristall ist ein *Nichtleiter*.
- Erst durch Energiezufuhr werden Elektronen aus den Gitterbindungen herausgelöst, so daß sie zur *Stromleitung* zur Verfügung stehen. Es kommt also auf den Energiezustand der Kristallelektronen an; das soll in einem späteren Abschnitt betrachtet werden.

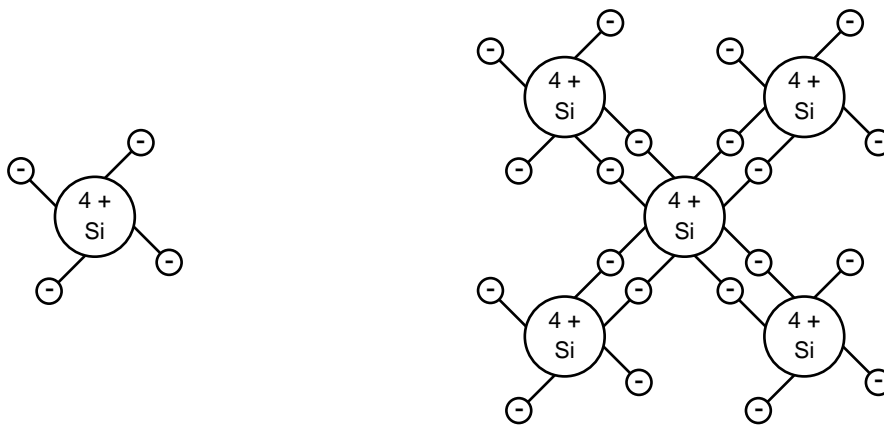


Bild 4.3: Schematische Darstellung eines idealen Siliziumkristalls
links: einzelnes Atom; rechts: ungestörtes Kristallgitter

4.1.2 Kristallgitterstörungen

Es wurde gezeigt, daß der ungestörte Halbleiterkristall ein Nichtleiter ist, da alle Valenzelektronen im Kristallgitter gebunden sind. Offenbar stehen erst dann Elektronen für eine Stromleitung zur Verfügung, wenn das ideale Kristallgitter *gestört* wird. Man kann, ausgehend von einem idealen Kristallgitter, durch den gezielten Einbau von Gitterstörungen der verschiedensten Art die Leitfähigkeit eines Halbleiters in weiten Grenzen verändern.

a) Punktförmige Gitterstörungen

Störungen 0. Ordnung sind punktförmige *Fremdatome*, die im Kristallgitter anstelle von Siliziumatomen eingebaut sind (Bild 4.4). Diese Art des Störstelleneinbaus wird technisch beherrscht und zur Herstellung von Siliziumkristallen mit gewünschten Leitungseigenschaften verwendet.

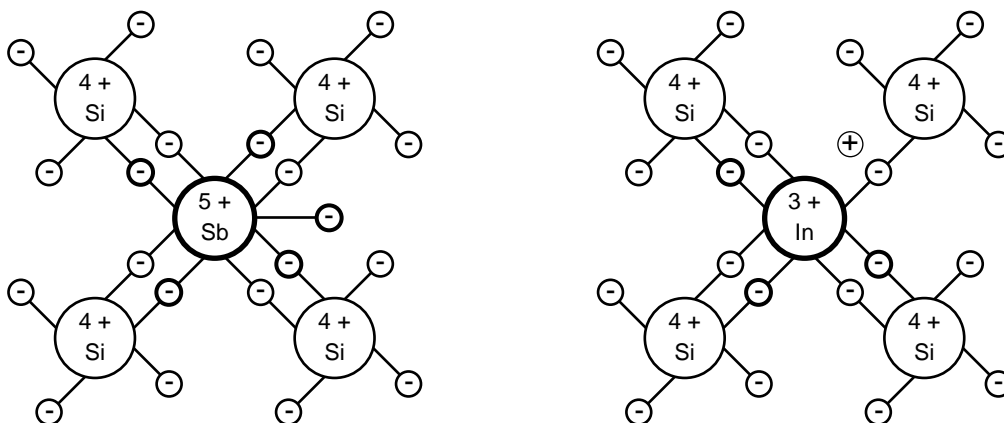


Bild 4.4: Schematische Darstellungen eines Siliziumkristalls mit eingebauten Störstellen
links: mit Donator-Atom; rechts: mit Akzeptor-Atom.

- Handelt es sich um 5-wertige Fremdatome - laut Periodischem System (Bild 4.1) zum Beispiel P, As oder Sb - so bleibt nach dem Absättigen der vier Gitterbindungen an jedem ersetzten Siliziumplatz je ein Valenzelektron übrig (Bild 4.4 links). Sie wirken als „Donatoren“, da sie ein Elektron zuviel mitbringen, das leicht abgelöst werden kann.

- Dagegen können 3-wertige Fremdatome - laut Periodischem System (Bild 4.1) zum Beispiel Al, Ga oder In - nicht alle vier Gitterbindungen an den Siliziumplätzen absättigen, die sie eingenommen haben; es bleibt je eine ungesättigte Gitterbindung übrig (Bild 4.4 rechts). Sie wirken als „Akzeptoren“, da die ungesättigte Gitterbindung leicht ein Elektron einfangen kann.

„Punktförmige“ Störstellen sind auch, so ungewohnt das klingen mag, in den Kristall eingestrahltes *Licht* und sogar die *Temperatur* des Kristalls. Am Beispiel des Elektrons haben wir gesehen, daß es je nach Bedarf als Korpuskel oder als Welle aufgefaßt werden kann. Entsprechend können Lichtwellen auch als Korpuskeln („Photonen“) betrachtet, die akustischen Gitterschwingungen des Kristalls, die mit steigender Temperatur zunehmen, ebenfalls als Korpuskeln („Phononen“) dargestellt werden. Sowohl Photonen als auch Phononen sind in der Lage, Kristallelektronen aus Gitterbindungen herauszuschlagen und sie für eine Stromleitung freizusetzen. Der nichtleitende Siliziumkristall ist also nur am absoluten Nullpunkt der Temperatur denkbar.

b) Linienförmige Gitterstörungen

Störungen 1. Ordnung sind die linienförmigen *Versetzungen*, die sich durch den Siliziumkristall ziehen können. Sie treten auf, wenn der Kristall plastisch verformt wird, sei es durch Biegen bei erhöhter Temperatur, sei es durch den Einbau zu vieler Fremdatome. Die Siliziumatome entlang einer Versetzung haben nur drei Nachbaratome, im Gegensatz zu den übrigen Kristallatomen, die deren vier haben. Sie können daher nur drei abgesättigte Gitterbindungen eingehen und haben eine freie Valenz übrig. Diese freie Valenz kann als überzähliges Valenzelektron oder als ungesättigte Gitterbindung aufgefaßt werden. Bei tiefen Temperaturen geben die Versetzungsatome die überschüssigen Valenzelektronen ab; sie wirken dann als Donatoren. Bei hohen Temperaturen fangen sie zur Absättigung der Gitterbindung ein zusätzliches Elektron ein; dann wirken sie als Akzeptoren. Wegen ihres veränderlichen Leitungstyps sucht man sie beim Herstellungsprozess zu vermeiden.

c) Flächige Gitterstörungen

Störungen 2. Ordnung sind die *Oberflächen* eines Halbleiterkristalls. Diese Störung eines regelmäßig sich fortsetzenden Kristallgitters ist stets vorhanden, da der Kristall ja irgendwo aufhören muß. Völlig reine Oberflächen hätten ebenfalls freie Valenzen, da den Oberflächenatomen ein oder mehrere der erforderlichen vier Nachbaratome fehlen. Beim Herstellungsprozess sind die Halbleiterstrukturen daher entweder im geschützten Kristallinnern anzuordnen oder die Oberflächen sind zu „versiegeln“, d.h. elektrisch zu neutralisieren, was bei Silizium (Si) relativ einfach mit einer Quarzglasschicht (SiO_2) erfolgen kann.

d) Räumliche Gitterstörungen

Störungen 3. Ordnung schließlich sind die räumlichen *Stapelfehler* des Kristalls, die sich während des Kristallwachstums bilden können und die ebenfalls störende elektrische Eigenschaften aufweisen. Auch sie sucht man beim Herstellungsprozess zu vermeiden.

4.1.3 Das Energiebändermodell

Das Bändermodell stellt die *Energiezustände* der Elektronen im Kristallgitter dar. Valenzelektronen, die in Gitterbindungen gebunden sind, befinden sich im „Valenzband“ auf dem Energieniveau E_V

(Bild 4.5). Im Fall des idealen, ungestörten Kristalls füllen alle Valenzelektronen das Valenzband voll aus. Das auf einem höheren Energieniveau E_C liegende „Leitungsband“, in dem sich Elektronen quasi frei bewegen können, ist dann leer. Weder ein vollbesetztes noch ein leeres Energieband tragen zur Leitfähigkeit etwas bei: Wie erwähnt ist der ungestörte Halbleiterkristall ein Nichtleiter. Zwischen den beiden „erlaubten“ Energiebändern liegt das sog. „verbotene Band“, das von Elektronen nur kurzzeitig bei einem Wechsel zwischen den erlaubten Bändern durchlaufen werden kann.

a) Eigenleitung

Erwärmt man den Halbleiterkristall so stark, daß genügend Phononen zur Überwindung der Energie

$$E_{CV} = E_C - E_V \quad 0 \rightarrow \oplus + \ominus \quad (4.1)$$

vorhanden sind, so sind diese in der Lage, Kristallelektronen aus den Gitterbindungen herauszuschlagen, das heißt, sie von der Oberkante E_V des Valenzbandes bis zur Unterkante E_C des Leitungsbandes anzuheben (Bild 4.5). Es entstehen dann sowohl Elektronen im Leitungsband als auch Defektelektronen im Valenzband, und zwar paarweise in genau gleicher Anzahl. Man spricht dann von der „Eigenleitung“, die vom Bandabstand E_{CV} des Kristalls und seiner Temperatur abhängt, nicht aber von einer Dotierung mit Fremdatomen. Ein eigenleitender Halbleiter ist schaltungstechnisch unbrauchbar, da sich seine Leitfähigkeit durch eine Dotierung nicht beeinflussen läßt. Die Eigenleitung wird in Silizium ab einer Temperatur von etwa 135°C dominierend, elektronische Bauelemente aus Silizium können deshalb nur unterhalb dieser Temperatur betrieben werden.

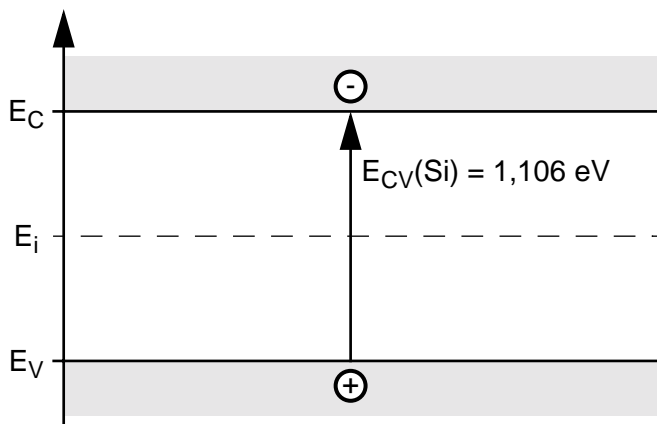


Bild 4.5: Eigenleitung durch Überwinden des „verbotenen Bandes“ der Breite E_{CV} .
 E_C - Energieniveau des Leitungsbandes; E_i - Inversionsenergie („Bandmitte“);
 E_V - Energieniveau des Valenzbandes.

b) n -Leitung mit Donatoren

Störstellen kann man grundsätzlich durch einen entsprechenden „Energieterm“ im Bändermodell darstellen.

Sind die Störstellen 5-wertige Fremdatome, so bleiben an ihrem jeweiligen Platz im Siliziumkristall nach dem Absättigen von vier Gitterbindungen ein Valenzelektron übrig, das man mit geringem Energieaufwand ablösen kann. Es genügt eine geringe Energiezufuhr

$$E_{CD} = E_C - E_D \qquad D^x \rightarrow D^+ + \ominus \qquad (4.2)$$

zum Beispiel durch Stöße von Phononen bei erhöhter Temperatur, um die überzähligen Elektronen von den Donatoratomen abzulösen und ins Leitungsband anzuheben, wo sie zur Stromleitung zur Verfügung stehen (Bild 4.6). *Man spricht dann von einem elektronen- oder n-leitenden Halbleiter.* Somit sind die zunächst neutralen Donatoren D^x im Bändermodell durch Energieterme bei E_D dicht unterhalb der Unterkante E_C des Leitungsbandes darzustellen. An den Störstellen bleiben 1-fach positiv geladene Donator-Ionen D^+ zurück, wie die obige Reaktionsgleichung zeigt.

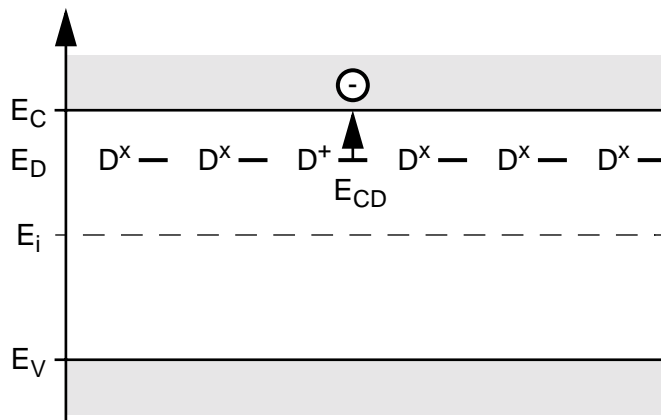


Bild 4.6: Energieterm E_D von Donatoren im Bändermodell

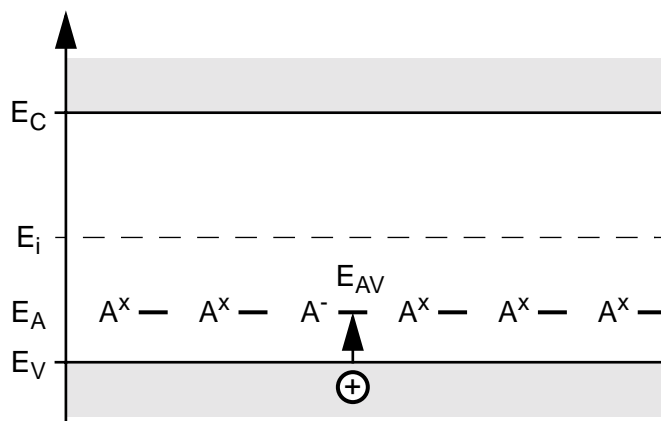


Bild 4.7: Energieterm E_A von Akzeptoren im Bändermodell

c) p -Leitung mit Akzeptoren

Sind die Störstellen 3-wertige Fremdatome, so können sie an ihrem jeweiligen Platz im Siliziumkristall nur drei Gitterbindungen absättigen. Die ungesättigte Gitterbindung jedes der zunächst neutralen Akzeptoren A^x sucht sich mit geringem Energieaufwand ein Elektron als zweiten Partner einzufangen. Dazu stehen genügend Elektronen im Valenzband zur Verfügung, die nur bis zu den entsprechenden Energietermen bei E_A dicht oberhalb der Oberkante E_V des Valenzbandes anzuheben sind (Bild 4.7), zum Beispiel wieder durch Phononenstöße der Energie

$$E_{AV} = E_A - E_V \qquad A^x \rightarrow A^- + \oplus \qquad (4.3)$$

wobei an den Störstellen 1-fach negativ geladene Akzeptor-Ionen A^- entstehen, wie die obige Reaktionsgleichung zeigt. Im Valenzband bleiben wegen der fehlenden Elektronen „Löcher“ zurück. Man kann diese Löcher durch positive Elementarladungen darstellen, die sich als quasifreie „Defektelektronen“ der Energie E_V im Valenzband fortbewegen und zur Stromleitung beitragen können. Man spricht dann von einem defektelektronen- oder *p-leitenden Halbleiter*. Man siedelt die Defektelektronen vor ihrer Freisetzung in den Energietermen E_A dicht oberhalb des Valenzbandes E_V an. Die Energiezufuhr E_{AV} befördert sie dann ins Valenzband.

d) Kompensation

Sind sowohl Donatoren als auch Akzeptoren in einem Siliziumkristall enthalten, so werden bei Energiezufuhr nicht Elektronen ins Leitungsband und/oder Defektelektronen ins Valenzband gebracht. Vielmehr tritt spontan, d.h. ohne Energiezufuhr von außen, die gegenseitige „Kompensation“ der Störstellen ein. Dabei wird Energie frei (Bild 4.8):

$$E_{DA} = E_D - E_A \qquad D^x + A^x \rightarrow D^+ + A^- \qquad (4.4)$$

Je nachdem, ob der Kristall nun mehr Donatoren oder mehr Akzeptoren enthält, läuft bei Energiezufuhr mit den überschüssigen Störatomen die Reaktion der Freisetzung von Ladungsträgern ab, so daß nur Elektronen oder nur Defektelektronen entstehen.

Die Leitfähigkeit eines kompensierten Halbleiters hängt demnach nicht von der Summe der Donator- und Akzeptoratom ab, sondern von deren *Differenz*. Hier deutet sich die Möglichkeit an, einen elektronenleitenden Halbleiter durch genügend starke Zufuhr von Akzeptoratom in einen defekt-leitenden Halbleiter „umzudotieren“ (bzw. umgekehrt).

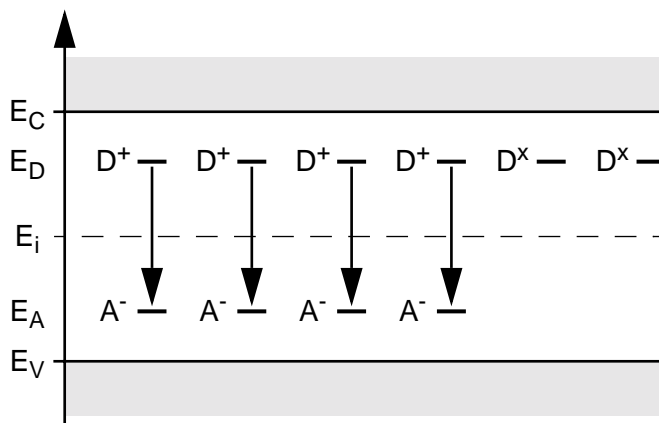


Bild 4.8: Kompensation von Akzeptoren durch Donatoren

4.2 Dichte der Ladungsträger

Zur rechnerischen Ermittlung der Anzahl der Ladungsträger in einem Halbleiterkristall, die zur Stromleitung zur Verfügung stehen, sind drei Beziehungen zu betrachten:

1. Das *Massenwirkungsgesetz*, das besagt, daß nach Erreichen eines thermischen Gleichgewichts zwar laufend Ladungsträger durch „Generation“ gebildet werden, daß aber ebenso viele durch „Rekombination“ wieder verschwinden.
2. Die *Störstellenbilanz*, die die neutralen und die ionisierten Störstellen im Kristall zusammenfaßt.
3. Die *Neutralitätsbedingung*, die ausdrückt, daß der Halbleiterkristall insgesamt stets elektrisch neutral bleibt.

Als Mengenangabe führt man die „Dichte“ der Ladungsträger ein, d.h. ihre Anzahl pro cm^3 : Sie sei n_n für Elektronen und n_p für Defektelektronen.

4.2.1 Die Eigenleitungsdichte

Ist die Temperatur des Halbleiterkristalls so hoch, daß genügend Phononen vorhanden sind, um so zahlreiche Elektronen aus dem Valenz- ins Leitungsband anzuheben, daß die Einflüsse anderer Kristallstörungen dagegen vernachlässigt werden können, so enthält der Kristall durch thermische Paarbildung („Generation“) gleich viele Elektronen wie Defektelektronen entsprechend der linken Reaktionsgleichung:



Es kommt auch laufend zur Vereinigung von Ladungsträgern („Rekombination“), weil Elektronen spontan aus dem Leitungs- ins Valenzband zurückfallen entsprechend der rechten Reaktionsgleichung. Da die Vereinigungsrate mit der Anzahl der Ladungsträger zunimmt, ist die Rekombinationsrate R zu den Dichten der beiden Rekombinationspartner proportional (Proportionalitätsfaktor sei ein Rekombinationskoeffizient r):

$$G = r n_n n_p \qquad R = r n_n n_p \qquad (4.5)$$

Im thermischen Gleichgewicht fallen gleich viele Elektronen aus dem Leitungs- ins Valenzband zurück wie durch Phononenstöße weiterhin laufend angehoben werden, d.h. die Generationsrate G ist gleich der Rekombinationsrate R , wie die Gleichungen (4.5) zeigen. Anstelle der Quotienten G/r $r = R/r$ führt man eine neue Größe „ n_i^2 “ ein, deren physikalische Bedeutung sogleich klar werden wird:

$$n_i^2 = n_n \cdot n_p \qquad (4.6)$$

Das ist das *Massenwirkungsgesetz* des störstellenfreien Halbleiters (für den eine *Störstellenbilanz* entfällt). So bleibt noch die *Neutralitätsbedingung* aufzustellen:

$$n_n = n_p \qquad (4.7)$$

woraus mit Gleichung (4.6) für den eigenleitenden Halbleiter folgt:

$$n_i = n_n = n_p$$

Die Größe n_i ist demnach die „Eigenleitungsdichte“, eine Materialkenngröße.

Die Festkörperphysik lehrt im einzelnen, daß die Eigenleitungsdichte n_i exponentiell vom Energieabstand E_{CV} zwischen Valenz- und Leitungsband und von der Temperatur T abhängt (wobei k die Boltzmannsche Konstante ist):

$$n_i \sim \exp(-E_{CV}/2kT) \quad \ln n_i = \text{const.} - E_{CV}/2kT \quad (4.8)$$

Das Logarithmieren beider Seiten der Gleichung (4.8) ermöglicht ihre graphische Darstellung durch eine einfache Gerade: Man trägt wie in Bild 4.9 auf der Ordinate $y = [\ln n_i]$ und auf der Abszisse $x = [1/kT]$ ab. Aus der Steigung dieser Geraden läßt sich der Bandabstand E_{CV} ermitteln:

$$y = \text{const.} - (E_{CV}/2) \cdot x \quad \frac{dy}{dx} = -E_{CV}/2$$

Für Silizium ergeben sich bei Zimmertemperatur die folgenden Zahlenwerte:

$$n_i(\text{Si}) = 1,5 \cdot 10^{10} \text{ cm}^{-3} \quad E_{CV}(\text{Si}) = 1,106 \text{ eV}$$

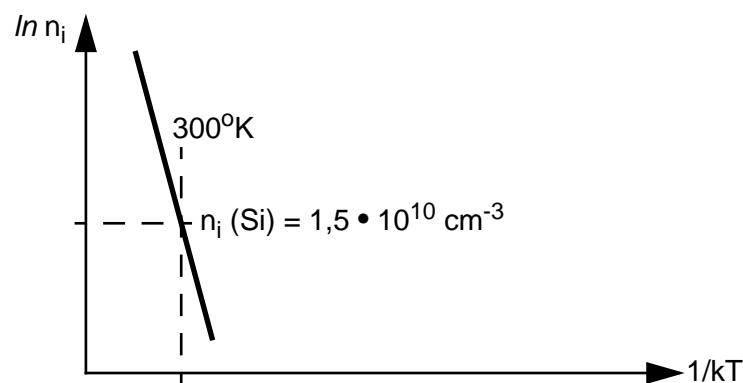


Bild 4.9: Temperaturverlauf der Eigenleitungsdichte n_i

Anzufügen ist noch die wichtige Tatsache, daß das Massenwirkungsgesetz:

$$n_i^2 = n_n \cdot n_p \quad (4.6)$$

nicht nur für eigenleitende Halbleiter gilt, sondern ganz allgemein auch für solche mit Störstellenleitung, d.h. mit einem deutlichen Überschuß an Elektronen oder Defektelektronen. Das bedeutet: Steigt durch geeignete Dotierung die Dichte n_n der Elektronen an, so muß die Dichte n_p der Defektelektronen im Halbleiterkristall entsprechend fallen, und umgekehrt. Man spricht daher bei n_i auch von der „Inversionsdichte“, da sie die Grenze zwischen Elektronen- und Defektelektronenleitung markiert; sie ist beim Umdotieren von Halbleiterkristallen eine wichtige Kenngröße.

Eine weitere Erkenntnis aus dem Massenwirkungsgesetz nach Gleichung (4.6) ist, daß in Halbleitern stets *beide* Ladungsträgerarten enthalten sind, wenn auch in der Regel in stark unterschiedlicher Dichte. Halbleiter enthalten folglich neben „Majoritätsträgern“ stets auch „Minoritätsträger“. Die Eigenleitung, bei der beide Ladungsträgerarten in gleicher Dichte vorhanden sind, ist ein Sonderfall.

4.2.2 Die Störleistungsichte

Um die elektrischen Eigenschaften höchstintegrierter Halbleiterschaltungen zu verstehen, ist es wichtig, die Temperaturabhängigkeiten der Ladungsträgerdichten zu kennen. Betrachtet man einen n-leitenden (p-leitenden) Halbleiterkristall, der als Störstellen zunächst neutrale Donatoren D^x (Akzeptoren A^x) enthalte, so gelten für die Generation von Ladungsträgern durch Ionisierung folgende Reaktionsgleichungen:



Bei der Abgabe der Ladungsträger entstehen ionisierte Störstellen D^+ (A^-), und zwar proportional zur Dichte n_{Dx} (n_{Ax}) der noch nicht ionisierten Störstellen:

$$G_D \sim n_{Dx} \quad G_A \sim n_{Ax} \quad (4.9)$$

Die Rekombination von Ladungsträgern mit ionisierten Störstellen D^+ (A^-) entsprechend der Reaktionsgleichungen:



ist proportional zu den Dichten der beiden Rekombinationspartner:

$$R_D \sim n_{D+} \cdot n_n \quad R_A \sim n_{A-} \cdot n_p \quad (4.10)$$

Da im thermischen Gleichgewicht die Generationsrate G gleich der Rekombinationsrate R ist, folgt durch Gleichsetzen der Gleichungen (4.9) und (4.10) das *Massenwirkungsgesetz* für einen mit Fremdatomen dotierten Halbleiterkristall, wobei K_D (K_A) Proportionalitätsfaktoren sind:

$$K_D \cdot n_{Dx} = n_{D+} \cdot n_n \quad K_A \cdot n_{Ax} = n_{A-} \cdot n_p \quad (4.11)$$

Die *Störstellenbilanz* ist rasch aufgestellt: Sie besagt, daß die Summe der neutralen und der ionisierten Störstellen gleich deren gesamter Anzahl sein muß:

$$n_D = n_{Dx} + n_{D+} \quad n_A = n_{Ax} + n_{A-} \quad (4.12)$$

Es fehlt noch die *Neutralitätsbedingung*, derzufolge die positiven und die negativen Ladungen sich gegenseitig kompensieren:

$$n_{D+} = n_n \quad n_{A-} = n_p \quad (4.13)$$

Aus den drei Gleichungen (4.11) bis (4.13) lassen sich die drei Unbekannten bestimmen: die Dichten der neutralen und der ionisierten Störstellen sowie die Dichte der Ladungsträger. Für letztere, die für die Stromleitung zur Verfügung stehen, ergibt sich:

$$n_n = K_D \left(-\frac{1}{2} + \sqrt{\frac{1}{4} + n_D / K_D} \right) \quad n_p = K_A \left(-\frac{1}{2} + \sqrt{\frac{1}{4} + n_A / K_A} \right) \quad (4.14)$$

Die Festkörperphysik gibt für die Massenwirkungskonstanten an:

$$K_D \sim \exp (- E_{CD} / kT) \quad K_A \sim \exp (- E_{AV} / kT) \quad (4.15)$$

Sie hängen von der Aktivierungsenergie E_{CD} (E_{AV}) der betreffenden Störstellen und von der Temperatur T ab.

a) Störstellenreserve

Nach Gleichung (4.15) sind die Massenwirkungskonstanten K_D (K_A) bei *tiefen* Temperaturen T klein gegenüber den Störstellendichten n_D (n_A), mit denen Halbleiter üblicherweise dotiert werden. Gleichung (4.14) wird dann näherungsweise:

$$\begin{aligned} K_D \ll n_D: & & K_A \ll n_A: \\ n_n = K_D(\sqrt{n_D/K_D}) = \sqrt{n_D K_D} & & n_p = K_A(\sqrt{n_A/K_A}) = \sqrt{n_A K_A} \end{aligned} \quad (4.16)$$

Setzt man in Gleichung (4.16) die Gleichung (4.15) ein, so erhält man:

$$n_n = \sqrt{n_D N_C / 2} \exp(-E_{CD} / 2kT) \quad n_p = \sqrt{n_D N_V / 2} \exp(-E_{AV} / 2kT) \quad (4.17)$$

Führt man wieder eine Umformung durch beidseitiges Logarithmieren durch:

$$\ln n_n = \frac{1}{2} \ln(n_D N_C / 2) - E_{CD} / 2kT \quad \ln n_p = \frac{1}{2} \ln((n_A N_V) / 2) - E_{AV} / 2kT$$

und trägt man wie in Bild 4.10 auf der Ordinate $y = [\ln n]$, auf der Abszisse $x = [1/kT]$ ab, so erhält man als graphische Darstellung eine Gerade, aus deren Steigung sich das Energieniveau E_{CD} (E_{AV}) der Störstellen ermitteln läßt, mit denen der Halbleiterkristall dotiert ist:

$$\begin{aligned} y &= \text{const}_c - (E_{CD} / 2) \cdot x & y &= \text{const}_v - (E_{AV} / 2) \cdot x \\ \frac{dy}{dx} &= -E_{CD} / 2 & \frac{dy}{dx} &= -E_{AV} / 2 \end{aligned}$$

Dieser bei *tiefen* Temperaturen zu beobachtende Temperaturverlauf (Bild 4.10) wird „Störstellenreserve“ genannt, da noch genügend nicht-ionisierte Störstellen in Reserve vorhanden sind, so daß mit steigender Temperatur immer mehr Ladungsträger erzeugt werden können.

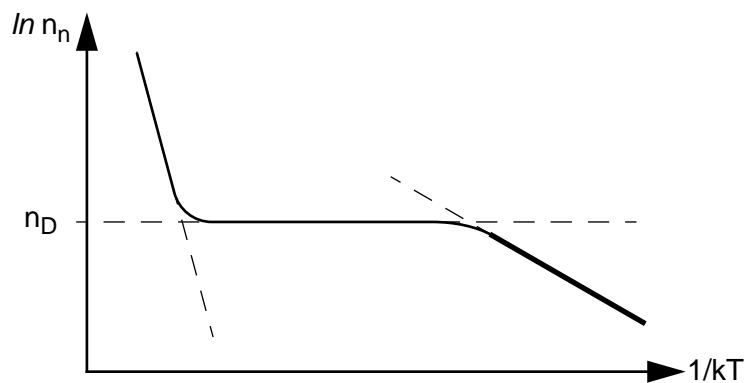


Bild 4.10: Temperaturverlauf der Ladungsträgerdichte; rechts: Störstellenreserve

b) Störstellenerschöpfung

Der Zustand der „Störstellenerschöpfung“ tritt bei *mittleren* Temperaturen ein, wenn die temperaturabhängigen Massenwirkungskonstanten K_D (K_A) groß gegenüber der jeweiligen Störstellendichte n_D (n_A) im Halbleiter werden. Dann wird die Gleichung (4.14) näherungsweise:

$$\begin{array}{ll} K_D \gg n_D: & K_A \gg n_A: \\ n_n = n_D & n_p = n_A \end{array} \quad (4.18)$$

Die Dichte der Ladungsträger ist dann gleich der Dichte der im Halbleiter enthaltenen Störstellen, da alle je einen Ladungsträger abgegeben haben. Die Menge der noch ionisierbaren Störstellen ist damit „erschöpft“ (Bild 4.11).

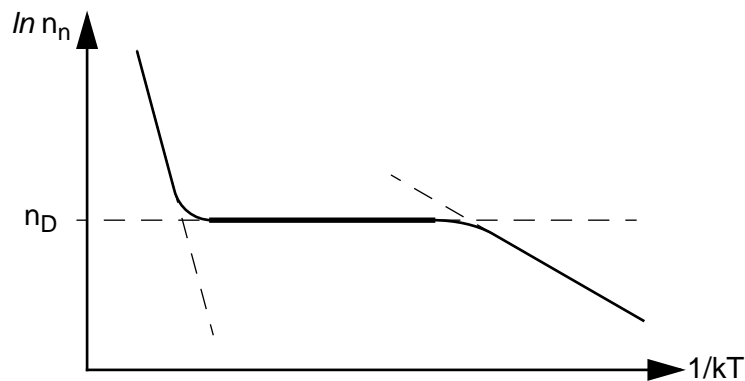
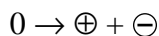


Bild 4.11: Temperaturverlauf der Ladungsträgerdichte; mitte: Störstellenerschöpfung

c) Eigenleitung

Bei *hohen* Temperaturen enthält der Halbleiterkristall schließlich so viele Phononen der Gesamtenergie E_{CV} , um eine beträchtliche Anzahl Elektronen von der Energie E_V des Valenzbandes direkt auf die Energie E_C des Leitungsbandes anzuheben, wobei im Valenzband durch die „Paarbildung“ gleich viele Defektelektronen zurückbleiben:



Ist die Temperatur des Halbleiterkristalls so hoch, daß die von den Störstellen abgegebenen Ladungsträger gegenüber den durch Paarbildung erzeugten Elektronen und Defektelektronen vernachlässigt werden können, so ist der Halbleiter weder überwiegend elektronen-, noch überwiegend defektelektronenleitend, sondern beides: Der Temperaturverlauf der Ladungsträgerdichte geht in die „Eigenleitung“ über (Bild 4.12).

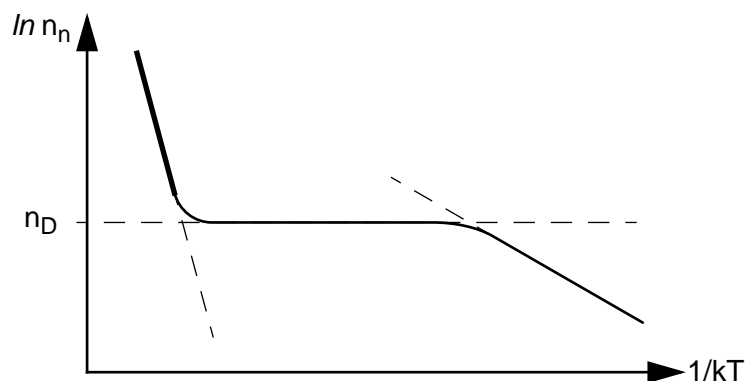


Bild 4.12: Temperaturverlauf der Ladungsträgerdichte; links: Eigenleitung

Passive Bauelemente, die elektrische Energie verbrauchen, können als Zweipole dargestellt werden. Aktive Bauelemente dagegen, die ein elektrisches Signal verstärken, besitzen *drei* äußere Anschlüsse: Eingangs- und Ausgangsanschluß sowie eine Steuerelektrode. Sie können als gesteuerter Schalter mit zwei unterschiedlichen Zuständen betrieben werden, einem stromleitenden und einem stromsperrenden. Sie sind daher für die Digitaltechnik und damit für die technische Informationsverarbeitung unverzichtbare Bauelemente. Aktive Bauelemente der Elektronik werden heute weitgehend aus dem Halbleitermaterial Silizium hergestellt. Die grundlegenden Schaltungsstrukturen sind der ins Kristallinnere eingebrachte PN-Übergang und der an der Kristalloberfläche gelegene MOS-Kondensator. Aus beiden lassen sich aktive, d.h. verstärkende Bauelemente aufbauen: aus ersterem der NPN-Transistor, aus letzterem der MOS-Transistor.

5.1 Der PN-Übergang

5.1.1 Das thermische Gleichgewicht

Man betrachte einen einheitlichen Siliziumkristall, der links mit Akzeptoren, rechts mit Donatoren dotiert sei (Bild 5.1 a). Bei Zimmertemperatur hat jedes Akzeptoratom genau ein Elektron aufgenommen, wodurch es negativ ionisiert wurde, während jedes Donatoratom genau ein Elektron abgegeben hat, so dass es positiv ionisiert ist. Im rechten Gebiet ist also ein Überschuß an freien Elektronen enthalten („N-Gebiet“), während im linken ein Mangel an freien Elektronen herrscht, was durch positiv geladene Defektelektronen dargestellt werden kann („P-Gebiet“). Die Akzeptor- und Donatorionen sind ortsfest im Kristallgitter eingebaut, während sich die genannten Ladungsträger quasi frei durch das Kristallgitter bewegen können. Da in jeweils beiden Gebieten die Anzahl der Ionen und die der Ladungsträger exakt gleich sind, ist der Siliziumkristall elektrisch neutral. Die Anzahl pro Volumeneinheit wird als „Dichte“ und mit dem Buchstaben n bezeichnet. Für die Akzeptorendichte n_A bzw. die Donatorendichte n_D gilt somit (Bild 5.1 b, c):

$$n_A = n_{pP} \quad n_D = n_{nN} \quad (5.1)$$

Die Defektelektronen im P-Gebiet mit der Dichte n_{pP} und die Elektronen im N-Gebiet mit der Dichte n_{nN} werden als *Majoritätsladungsträger* bezeichnet. Im thermischen Gleichgewicht gilt das sog. Massenwirkungsgesetz:

$$n_n \cdot n_p = n_i^2 \quad n_i = 1,5 \cdot 10^{10} \text{ cm}^{-3} \quad (5.2)$$

Dabei ist n_i die „Eigenleitungsdichte“ des Siliziums; angegeben ist der bei 300°K („Zimmertemperatur“) gemessene Wert. Danach enthält sowohl das P- als auch das N-Gebiet grundsätzlich auch *Minoritätsladungsträger*: das P-Gebiet auch Elektronen der Dichte n_{nP} , das N-Gebiet auch Defektelektronen der Dichte n_{pN} (Bild 5.1 c):

$$n_{nP} \cdot n_{pP} = n_i^2 \quad n_{nN} \cdot n_{pN} = n_i^2 \quad (5.3)$$

Die Majoritätsladungsträger bleiben, da sie quasi frei beweglich sind, nicht in ihren Gebieten, sondern sie dringen auch ins jeweils benachbarte Gebiet ein, wo sie Minoritätsladungsträger darstellen: Die Defektelektronen fließen nach rechts ins N-, die Elektronen nach links ins P-Gebiet. Dieser „Diffusionsvorgang“ ist analog zum Bestreben von Gasmolekülen, stets den gesamten zur Verfügung stehenden Raum gleichmäßig auszufüllen, und er ist charakteristisch für Halbleiter.

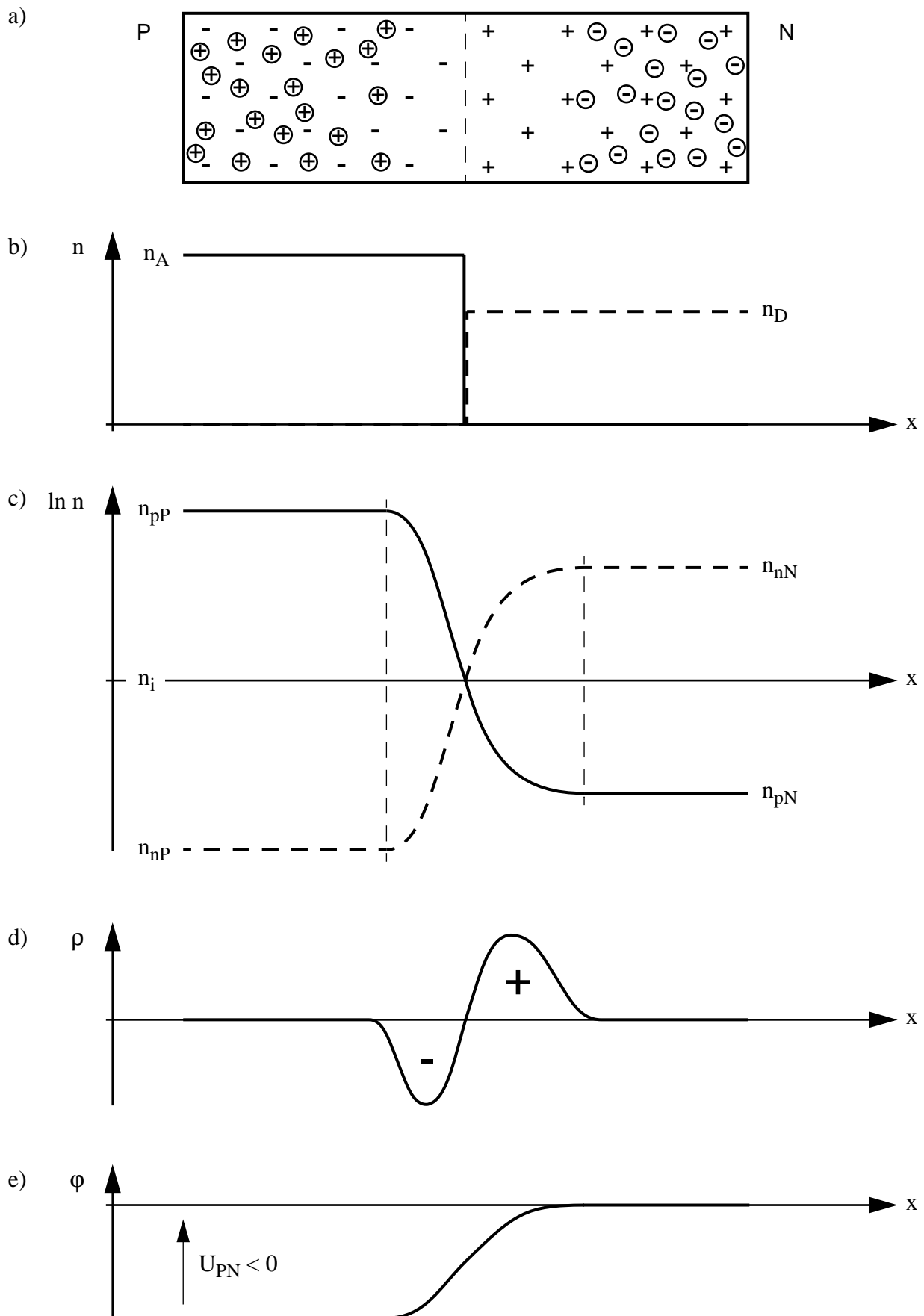


Bild 5.1: PN-Übergang im thermischen Gleichgewicht (Erläuterung im Text)

- Der entstehende *Diffusionsstrom* fließt ohne angelegte Spannung, gehorcht also nicht dem Ohmschen Gesetz. Er fließt in Richtung eines *Dichtegefälles* der Ladungsträger, d.h. vom Gebiet hoher zum Gebiet niedriger Dichte.
- Der vertrautere *Feldstrom* fließt bekanntlich in Richtung eines *Potentialgefälles* und gehorcht damit dem Ohmschen Gesetz.

Das Übergangsgebiet zwischen P- und N-Gebiet verarmt durch den Diffusionsstrom an Majoritätsladungsträgern (Bild 5.1 c). Auf beiden Seiten des PN-Übergangs entsteht eine hochohmige Schicht („Sperrschicht“). Aber auch in der Sperrschicht gilt das in Gl.(5.2) angegebene Massenwirkungsgesetz, d.h. direkt am PN-Übergang gilt:

$$n_n = n_p = n_i \quad (5.4)$$

Neben den quasi freien Ladungsträgern sind auch die fest im Kristallgitter verankerten Dotierungen zu betrachten, deren Ladung von den von ihnen abgegebenen Majoritätsladungsträgern grundsätzlich kompensiert wird. Da in der Sperrschicht aber aufgrund des Diffusionsvorgangs Majoritätsladungsträger fehlen, stellen die zurückbleibenden Dotierungen dort eine Raumladung ρ dar (Bild 5.1 d), die ein elektrisches Feld und damit eine Potentialdifferenz ϕ zwischen P- und N-Gebiet zur Folge hat (Bild 5.1 e). Das entstehende Potentialgefälle ist so gepolt, dass es die wegdiffundierenden Ladungsträger in ihre jeweiligen Stammgebiete zurücktreibt. Durch das Potentialgefälle entsteht ein Feldstrom, der dem oben erwähnten Diffusionsstrom genau entgegengerichtet ist. Es stellt sich bei nach außen hin stromlosem PN-Übergang ein Gleichgewichtszustand ein, bei dem die Feldstromstärke gleich der Diffusionsstromstärke ist. Den Potentialunterschied, der den Dichteunterschied kompensiert, nennt man die „Diffusionsspannung“ U_{PN} .

5.1.2 Durchlaß- und Sperrbereich

Legt man eine äußere Spannungsquelle so an, dass ihr positiver Pol am P-Gebiet, ihr negativer am N-Gebiet anliegt, so werden die Majoritäten durch die Sperrschicht in Richtung des Nachbargebietes getrieben, der Diffusionsstrom wird größer als der Feldstrom und auch im äußeren Stromkreis fließt ein meßbarer Strom. Das Potentialgefälle zwischen P- und N-Gebiet wird um den Wert der angelegten Spannung erhöht, die Breite der Sperrschicht nimmt ab. Da sie nicht kleiner als Null werden kann, kann auch die Potentialdifferenz nicht kleiner als Null werden. Das bedeutet, dass eine Spannung, die größer ist als die Diffusionsspannung (bei Silizium etwa 1 Volt), an den PN-Übergang nicht angelegt werden kann, da sonst der Strom über alle Grenzen wachsen müsste: Der PN-Übergang ist jetzt in *Durchlaßrichtung* gepolt.

Dreht man die Polarität der äußeren Spannungsquelle um, so dass ihr negativer Pol am P-Gebiet, ihr positiver am N-Gebiet anliegt, so werden die Majoritäten nach außen abgesaugt, der Feldstrom wird größer als der Diffusionsstrom, die Sperrschicht verbreitert sich, die Diffusionsspannung nimmt um den Wert der angelegten Spannung zu. Es kann nur noch ein kleiner Strom über den PN-Übergang fließen, der von den immer vorhandenen Minoritätsladungsträgern herrührt: Der PN-Übergang ist jetzt in *Sperrichtung* gepolt.

Bild 5.2 zeigt die Durchlaßkennlinie $I = f(U)$ der Siliziumdiode. Man beachte die charakteristische „Schleusenspannung“ $U_S \approx 0,7$ Volt, ab der erst ein merklicher Durchlaßstrom fließt.

Zusammenfassend kann man sagen, dass der PN-Übergang zwei charakteristische Eigenschaften aufweist:

- **Gleichrichterwirkung:** Im stromlosen Zustand herrscht am PN-Übergang ein Gleichgewicht zwischen Diffusions- und Feldströmen beider Ladungsträgerarten aufgrund der sich aufbauenden Diffusionsspannung zwischen P- und N-Gebiet. Eine von außen angelegte Spannung stört dieses Gleichgewicht. Je nach deren Polarität ergibt sich entweder ein großer Durchlaß-, oder ein kleiner Sperrstrom.

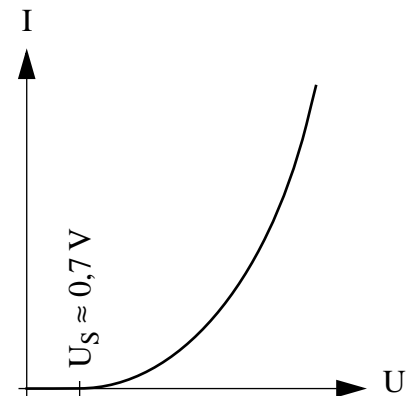


Bild 5.2: Durchlaßkennlinie der Siliziumdiode

- **Sperrschichtkapazität:** Da die Sperrschicht an beweglichen Minoritätsträgern verarmt, entsteht dort eine Raumladung aus nicht kompensierten Dotierungsionen, die ein elektrisches Feld und eine Potentialdifferenz zwischen P- und N-Gebiet zur Folge hat. Der PN-Übergang stellt somit im Prinzip einen geladenen Kondensator dar.

5.2 Bipolare Transistoren

Wir wollen uns hier auf den am häufigsten verwendeten NPN-Typ beschränken. Der dazu komplementäre PNP-Typ unterscheidet sich nur in der Polarität der Ladungsträger, der Betriebs- und der Ansteuerspannungen. Die Bezeichnung „bipolar“ soll ausdrücken, dass seine Wirkungsweise auf dem Zusammenwirken beider Ladungsträgertypen, der Majoritäten und der Minoritäten beruht.

5.2.1 Aufbau und Wirkungsweise

Ein bipolarer Transistor besteht aus zwei eng benachbarten PN-Übergängen. Bei einer NPN-Struktur enthalten die beiden äußeren N-Gebiete einen Elektronenüberschuß, das mittlere P-Gebiet einen Überschuß an Defektelektronen (Bild 5.3). Im aktiven Betrieb wird einer der beiden PN-Übergänge in Durchlaß-, der andere in Sperrichtung gepolt. Man nennt den leitenden PN-Übergang den *Basis-Emitter-Übergang*; denn aufgrund des zugunsten des Diffusionsstroms verschobenen Stromgleichgewichts *sendet* das N-Emittergebiet seine Elektronen in erhöhter Dichte ins P-Basisgebiet (wo sie Minoritätsträger darstellen). Ein Teil der ins Basisgebiet injizierten Elektronen rekombiniert mit den dort vorhandenen Majoritätsträgern, den Defektelektronen. Dieser Ladungsverlust muss von außen aus der Spannungsquelle U_{BE} , die den Basis-Emitter-Übergang in Durchlaßrichtung polt, als Basisstrom I_B nachgeliefert werden (Bild 5.4).

Ist die Lebensdauer der Minoritätsträger in der P-Basis, hier der Elektronen, relativ groß und die Basisschicht entsprechend dünn, so gelangen relativ viele der vom N-Emitter in die P-Basis injizierten Elektronen bis zum gesperrten PN-Übergang, der jedoch nur für die Majoritätsladungsträger in Sperrichtung gepolt ist. Für Minoritätsladungsträger, d.h. hier für die das P-Basisgebiet durchlaufenden Elektronen, stellt diese Polung die Durchlaßrichtung dar, so dass die injizierten Elektronen bis ins dritte, ein N-Gebiet hinein laufen, wo sie als Majoritäten wieder *gesammelt* werden. Man nennt daher den gesperrten PN-Übergang den *Basis-Collector-Übergang*.

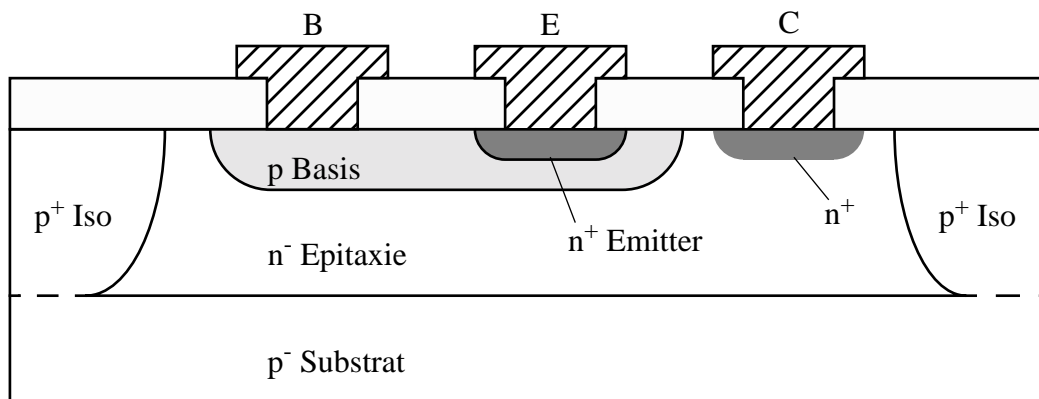


Bild 5.3: Schematischer Aufbau eines bipolaren Transistors vom NPN-Typ (Querschnitt)

Die Überschwemmung des in Sperrrichtung gepolten Basis-Collector-Übergangs mit Elektronen erhöht dessen Sperrstrom beträchtlich. Die Elektronen werden von der Spannungsquelle U_{CE} abgesaugt, was einem Collectorstrom I_C entspricht (Bild 5.4). Dieselbe Spannungsquelle liefert auch die vom Emitter injizierten Elektronen als Emitterstrom I_E nach. (Dabei ist zu beachten, dass die technische Stromrichtung definitionsgemäß entgegen der Flußrichtung der Elektronen angenommen wird.) Die Wirkungsweise des bipolaren Transistors beruht darauf, dass eine relativ niedrige Basis-Emitter-Durchlaßspannung $U_{BE} \approx 0,7 \text{ V}$, d.h. ein kleiner Basisstrom I_B für eine starke Injektion von Ladungsträgern ins Basisgebiet sorgt, die von der Ausgangsspannung $U_{CE} \approx 5,0 \text{ V}$ als verstärkter Collectorstrom I_C abgesaugt und als annähernd gleich hoher Emitterstrom I_E nachgeliefert werden. Es läßt sich mit einer niedrigen Eingangsleistung eine hohe Ausgangsleistung steuern: Der Transistor dient der *Leistungsverstärkung*.

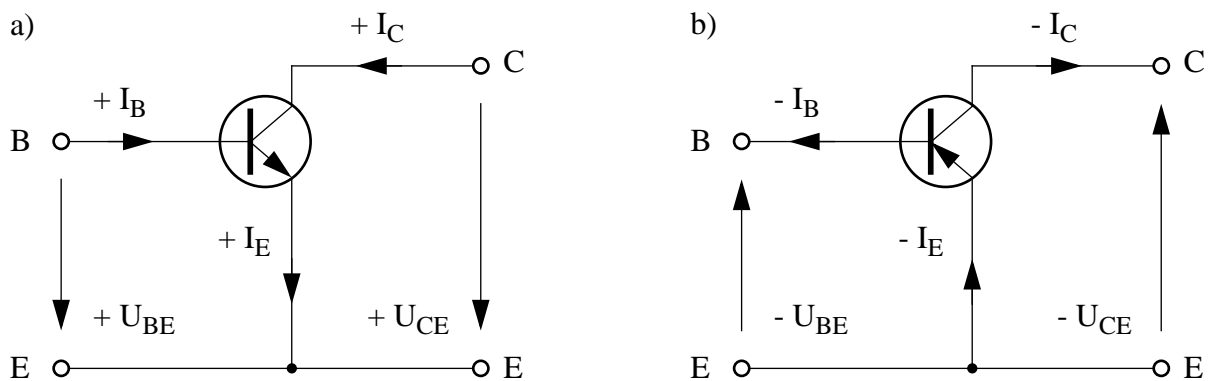


Bild 5.4: Schaltsymbole bipolarer Transistoren; B: Basis, E: Emitter, C: Collector.

a) NPN-Typ

b) PNP-Typ

Formal läßt sich ein aktives Bauelement wie der Transistor, der über drei Anschlüsse verfügt, auch als Vierpol darstellen, der einen Eingangs- mit einem Ausgangstromkreis koppelt (Bild 5.4). Im Eingangskreis wirken die Basis-Emitter-Spannung U_{BE} und der Basisstrom I_B , im Ausgangskreis der Collectorstrom I_C und die Ausgangsspannung U_{CE} . Man definiert folgende Verstärkungsfaktoren, wobei es sich um Gleichstromverstärkungen handelt:

$$\text{Alpha} = I_C : I_E = 0,90 \dots 0,99$$

$$\text{Beta} = I_C : I_B = 10 \dots 100 \quad (5.5)$$

5.2.2 Kennlinien und Betriebsbereiche

Eine Kennlinie zeigt den Zusammenhang zwischen zwei Größen, hier zwischen Strom und Spannung. Führt man eine dritte Größe ein („Parameter“), so ergibt sich für unterschiedliche, konstante Parameterwerte ein Kennlinienfeld. Beim bipolaren Transistor unterscheidet man drei Kennlinien:

- Die Steuerkennlinie $I_C = f(U_{BE})$ für $U_{CE} = \text{const.}$ (Bild 5.5 a). Sie veranschaulicht den Durchgriff der Eingangs- auf die Ausgangsgrößen. Ist die Eingangsspannung größer als die Schleusenspannung: $U_{BE} > U_S$, so werden vom Emitter Elektronen in die Basiszone injiziert, die dann von der Ausgangsspannung $U_{CE} > 0$ zum Collector hin abgesaugt werden und den Ausgangsstrom I_C bilden.
- Die Eingangskennlinie $I_B = f(U_{BE})$ für $U_{CE} = \text{const.}$ (Bild 5.5 b). Sie veranschaulicht die Belastung der ansteuernden Signalquelle. Durch eine Eingangsspannung $U_{BE} > U_S$ werden Elektronen in die Basiszone injiziert, die zum Teil mit den dort vorhandenen Defektelektronen rekombinieren, was sich in einem kleinen Basisstrom I_B äußert. Bei kurzgeschlossenem Eingang $U_{BE} = 0$ fließt der Sperrstrom $I_B = -I_S$ der Basis-Collector-Diode.
- Das Ausgangskennlinienfeld $I_C = f(U_{CE})$ für $(U_{BE} - U_S) = \text{const.}$ als Parameter (Bild 5.6). Bei genügend hoher Ausgangsspannung U_{CE} wirkt der Transistor als Stromquelle $I_C = \text{const.}$, was zur Signalverstärkung verwendet werden kann.

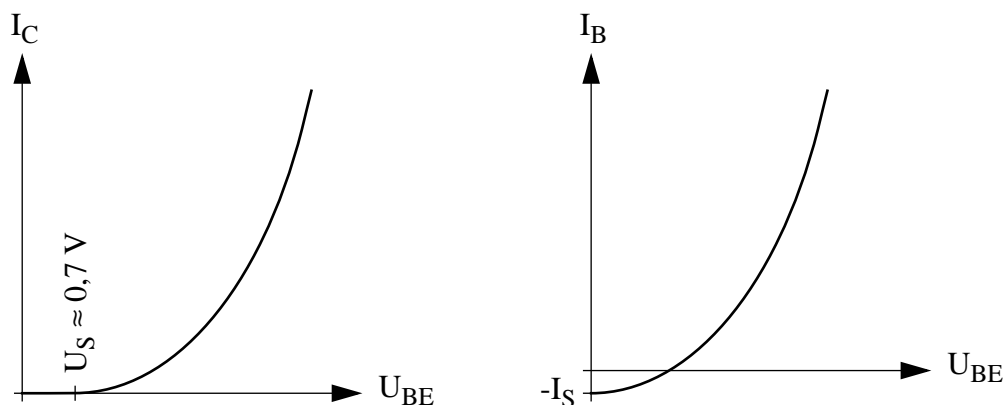


Bild 5.5: Steuer- und Eingangskennlinie eines bipolaren NPN-Transistors

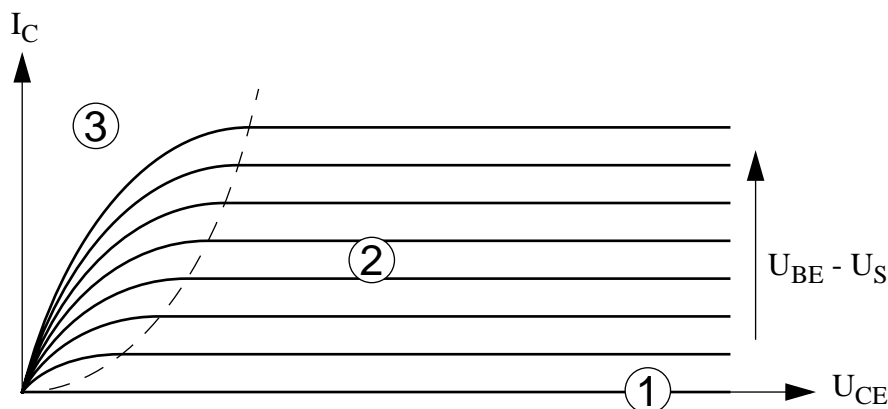


Bild 5.6: Ausgangskennlinienfeld eines bipolaren NPN-Transistors;
1 - gesperrter Bereich, 2 - linearer Bereich, 3 - gesättigter Bereich

Je nach angelegten Spannungswerten unterscheidet man beim bipolaren Transistor drei Betriebsbereiche, die im Ausgangskennlinienfeld nach Bild 5.6 angegeben sind:

- **Gesperrter Bereich:** Beide PN-Übergänge sind gesperrt ($U_{BE} \leq U_S$; $U_{BC} \leq U_S$).
Dann fließen über den Transistor trotz angelegter Ausgangsspannung $U_{CE} > 0$ nur extrem kleine Restströme. Der Transistor entspricht einem hohen Widerstandswert, d.h. einem „geöffneten Schalter“.
- **Linearer Bereich:** Ein PN-Übergang ist leitend, einer gesperrt ($U_{BE} > U_S$; $U_{BC} \leq U_S$).
Die Injektion von Elektronen in die Basiszone wird durch die Eingangsspannung U_{BE} bestimmt, die hier konstant gehalten wird; der Collectorstrom I_C hängt kaum von der Ausgangsspannung U_{CE} ab. In der Digitaltechnik wird dieser Bereich möglichst schnell durchlaufen, in der Verstärkertechnik ist es der „Normalbetrieb“.
- **Gesättigter Bereich:** Beide PN-Übergänge sind leitend ($U_{BE} > U_S$; $U_{BC} > U_S$).
Verringert man die Ausgangsspannung auf $U_{CE} < U_{BE}$, so wird $U_{BC} > 0$, d.h. auch der Basis-Collector-Übergang wird in Durchlaßrichtung gepolt. Dann injizieren sowohl Emitter als auch Collector in entgegengesetzten Richtungen Ladungsträger in die Basiszone, wodurch Collector- und Basisstrom stark ansteigen; zwischen Collector und Emitter fällt nur noch eine geringe Restspannung ab. Der Transistor entspricht einem niedrigen Widerstandswert, d.h. einem „geschlossenen Schalter“.

Schließlich kann man beim NPN-Transistor die Funktionen des N-Emitters und des N-Collectors auch vertauschen, indem man die angelegte Versorgungsspannung U_{CE} umpolt und den Transistor „invers“ betreibt. Auch in diesem Betriebszustand lassen sich die genannten drei Betriebsbereiche unterscheiden. Bipolare Transistoren sind jedoch in der Regel stark unsymmetrisch aufgebaut (vgl. Bild 5.3), so dass die Stromverstärkung im Inversbetrieb wesentlich kleiner ist als im Normalbetrieb.

5.3 Der MOS-Kondensator

Die Oberfläche eines Siliziumkristalls (Si) kann relativ einfach oxidiert werden, wobei sie mit einer dünnen Schicht Siliziumdioxid (SiO_2) überzogen wird, chemisch widerstandsfähiges und elektrisch fast perfekt isolierendes Quarzglas. Wird darauf eine dünne Schicht Metall aufgedampft, so erhält man eine Metall-Oxid-Silizium-Schichtstruktur: einen MOS-Kondensator. Es ist zu beachten, dass der MOS-Kondensator auch ohne angelegte Spannung bereits dreierlei Ladungen enthält:

- Q_{ox} : Die isolierende Oxidschicht enthält stets Natriumionen, d.h. positive Ladungen. Während des Oxidationsprozesses läßt sich deren Einbau grundsätzlich nicht vermeiden.
- Q_{ss} : An der Grenzfläche zwischen Silizium und Oxidschicht befinden sich unabgesättigte Gitterbindungen, die Elektronen abgeben oder einfangen können („surface states“). Man beobachtet bei Si, das mit SiO_2 überzogen ist, dass sich die Oberflächenzustände jeweils wie die Dotierungsatome im darunter befindlichen Siliziumkristall verhalten.
- Q_{Si} : Die ionisierten Störstellen im Siliziumkristall, wobei es sich um Donatoren oder Akzeptoren handeln kann, stellen eine fest im Kristallgitter verankerte Raumladung dar.

Während die Oxidladung Q_{ox} stets positiv ist, hängt das Vorzeichen der beiden anderen Ladungen offenbar von der Dotierung des Siliziumkristalls ab. Sie können sich deshalb in ihrer Wirkung addieren oder subtrahieren. Man unterscheidet daher zwei Typen von MOS-Kondensatoren.

- **Anreicherungstyp:** Eine von außen angelegte Spannung U_{MOS} erzeugt an der Oberfläche des Siliziumkristalls einen leitenden Kanal, d.h. an der Si-SiO₂-Grenzschicht werden Ladungsträger *angereichert*. Ohne angelegte Spannung ist die MOS-Struktur „selbstsperrend“.
- **Verarmungstyp:** Eine von außen angelegte Spannung U_{MOS} beseitigt einen spontan an der Oberfläche des Siliziumkristalls vorhandenen leitenden Kanal, d.h. die Si-SiO₂-Grenzschicht *verarmt* an Ladungsträgern. Ohne angelegte Spannung ist die MOS-Struktur „selbstleitend“.

5.3.1 Anreicherungstyp

Eine an den MOS-Kondensator angelegte Spannung muss beim Anreicherungstyp eine Schwellspannung U_{th} („threshold voltage“) betragsmäßig überschreiten, damit ein leitender Kanal *entsteht*. Ist der MOS-Kondensator mit N-leitendem Silizium aufgebaut (Bild 5.7), so gilt für die vorhandene Gesamtladung:

$$Q_{ges} = Q_{ox} + Q_{ss} + Q_{Si} > 0 \quad (5.6)$$

Da diese Ladungen durch die von außen angelegte Spannung U_{MOS} kompensiert werden müssen, bevor ein Kanal erzeugt werden kann, gilt für die Schwellspannung in diesem Fall:

$$U_{th} = -U_{ox} - U_{ss} - U_{Si} < 0 \quad (5.7)$$

Ist $U_{MOS} \leq U_{th}$, so werden durch die negativ geladene Metallschicht des MOS-Kondensators an der Si-Oberfläche Defektelektronen angereichert: es entsteht ein *P-leitender Kanal*.

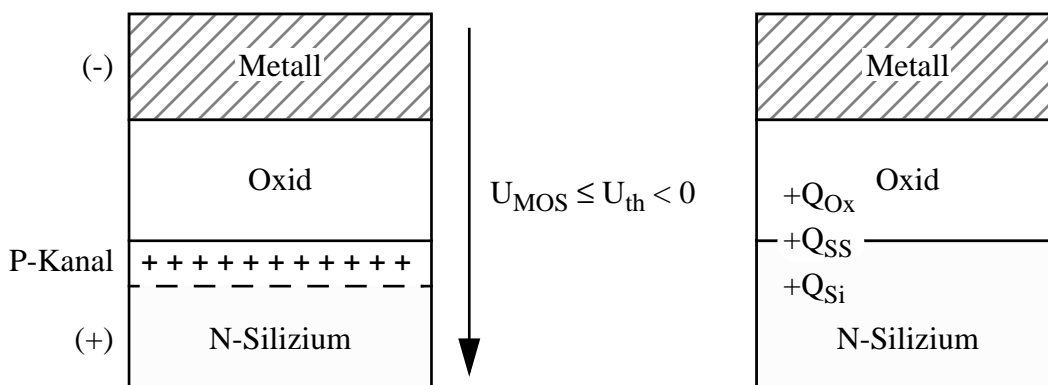


Bild 5.7: P-leitender Kanal vom Anreicherungstyp auf N-leitendem Silizium

Ist der MOS-Kondensator mit P-leitendem Silizium aufgebaut (Bild 5.8), so gilt für die wirksame Gesamtladung, wobei für die Ladungsverhältnisse $Q_{ox} < Q_{ss} + Q_{Si}$ angenommen wird:

$$Q_{ges} = Q_{ox} - Q_{ss} - Q_{Si} < 0 \quad (5.8)$$

Da die Ladungen durch die angelegte Spannung U_{MOS} kompensiert werden müssen, bevor ein Kanal erzeugt werden kann, gilt für die Schwellspannung:

$$U_{th} = -U_{ox} + U_{ss} + U_{Si} > 0 \quad (5.9)$$

Bei $U_{\text{MOS}} \geq U_{\text{th}}$ werden durch die positiv geladene Metallschicht des MOS-Kondensators an der Si-Oberfläche Elektronen angereichert: es entsteht ein *N-leitender Kanal*.

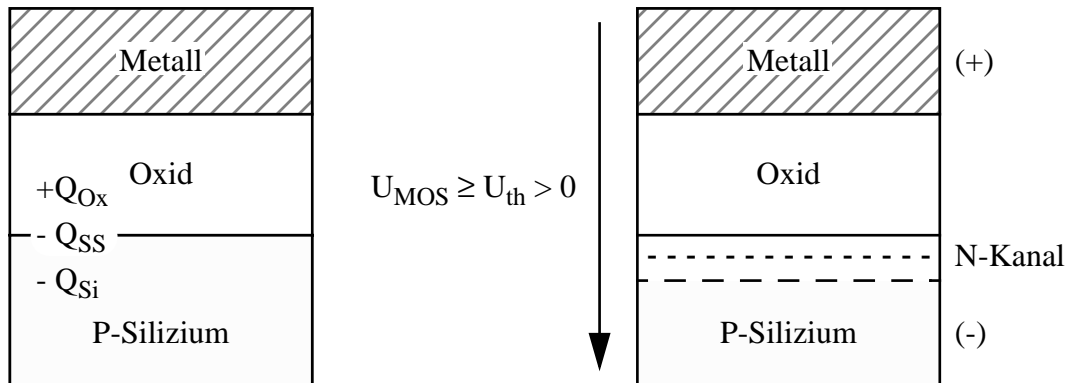


Bild 5.8: N-leitender Kanal vom Anreicherungstyp auf P-leitendem Silizium

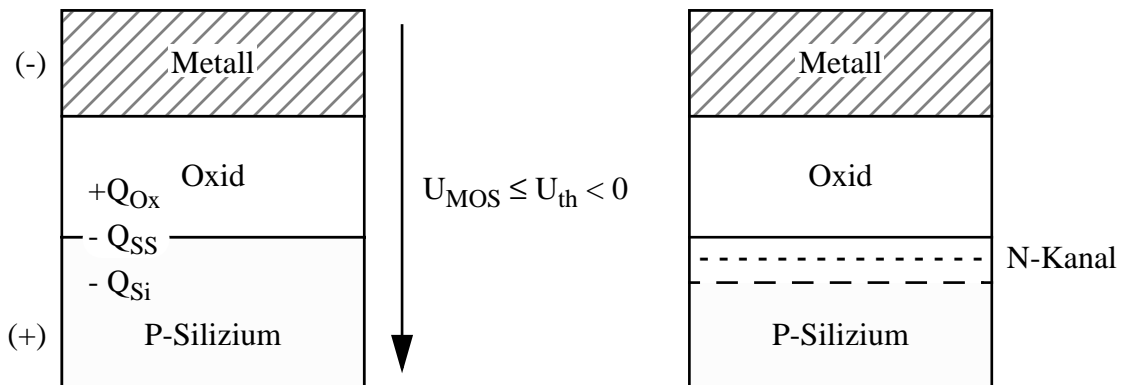


Bild 5.9: N-leitender Kanal vom Verarmungstyp auf P-leitendem Silizium

5.3.2 Verarmungstyp

Die angelegte Spannung muss beim Verarmungstyp die Schwellspannung U_{th} ("threshold voltage") betragsmäßig überschreiten, damit der leitende Kanal *verschwindet*. Ist der MOS-Kondensator mit P-leitendem Silizium aufgebaut (Bild 5.9), so gilt für die Gesamtladung wieder Gl.(5.8), wobei für die Ladungsverhältnisse in diesem Fall aber $Q_{\text{ox}} > Q_{\text{ss}} + Q_{\text{Si}}$ angenommen wird:

$$Q_{\text{ges}} = Q_{\text{ox}} - Q_{\text{ss}} - Q_{\text{Si}} > 0$$

Bei den angenommenen Ladungsverhältnissen gilt für die Schwellspannung:

$$U_{\text{th}} = -U_{\text{ox}} + U_{\text{ss}} + U_{\text{Si}} < 0 \quad (5.10)$$

Ist $U_{\text{MOS}} \leq U_{\text{th}}$, so werden durch die negativ geladene Metallschicht des MOS-Kondensators die Elektronen, die bereits ohne angelegte Spannung dort vorhanden waren und einen *N-leitenden Kanal* gebildet hatten, verdrängt und der Kanalbereich verarmt an Ladungsträgern.

Ist der MOS-Kondensator mit N-leitendem Silizium aufgebaut, so gilt für die Ladungen wieder Gl. (5.6) und für die Schwellspannung Gl.(5.7):

$$Q_{\text{ges}} = Q_{\text{ox}} + Q_{\text{ss}} + Q_{\text{Si}} > 0 \quad U_{\text{th}} = -U_{\text{ox}} - U_{\text{ss}} - U_{\text{Si}} < 0$$

Hier ist die Schwellspannung immer $U_{th} < 0$. Ein P-leitender Kanal vom Verarmungstyp würde aber $U_{th} > 0$ erfordern, was gleichbedeutend mit dem Einbau negativer Ionen in der Oxidschicht wäre. Da Siliziumdioxid aber stets einen Überschuß an positiv geladenen Natriumionen enthält, kommt der P-leitende Kanal vom Verarmungstyp in der Praxis nicht vor.

5.4 MOS-Transistoren

Wir wollen uns hier auf die Betrachtung des MOS-Transistors mit N-Kanal vom Anreicherungstyp beschränken; der P-Kanal-Typ unterscheidet sich nur in den Polaritäten der Ladungsträger und der angelegten Spannungen. Man findet die beiden zueinander komplementären Kanaltypen in der CMOS-Schaltungstechnik („Complementary Metal-Oxide-Silicon“). Die Bezeichnung „unipolar“ soll ausdrücken, dass die Wirkungsweise auf dem Einfluß eines elektrischen Feldes auf die Majoritätsträger beruht, während die Minoritätsträger hier keine Rolle spielen. MOS-Transistoren vom Verarmungstyp finden nur als passive, zum Zweipol verdrahtete Bauelemente Verwendung.

5.4.1 Aufbau und Wirkungsweise

Grundsätzlich besteht ein MOS-Transistor aus einem MOS-Kondensator, dessen an der Oberfläche des Siliziumkristalls befindlicher, schaltbarer Kanal zwei gesperrte PN-Übergänge bei Bedarf miteinander leitend verbindet (Bild 5.10). Die Leitfähigkeit des Kanals kann durch ein senkrecht dazu wirkendes elektrisches Feld gesteuert werden; man bezeichnet ihn deshalb auch als Feldeffekttransistor. Der selbstsperrende Anreicherungstyp eignet sich besonders gut für die Höchstintegration, die Millionen von Bauelementen auf einem einzigen Siliziumchip von etwa einem Quadratzentimeter Fläche unterbringt.

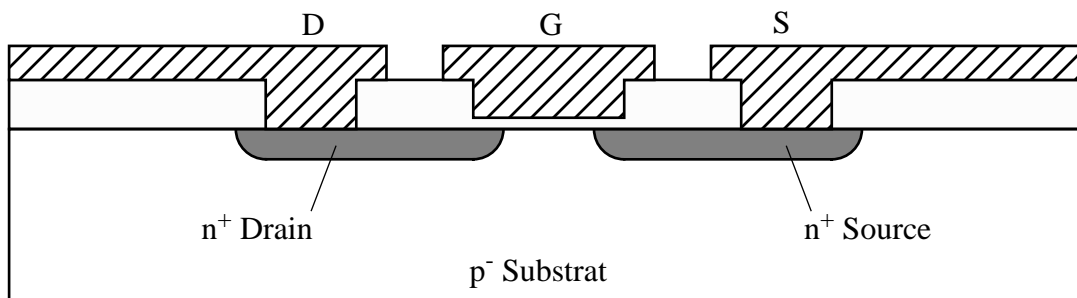


Bild 5.10: Schematischer Aufbau eines MOS-Transistors mit N-Kanal (Querschnitt)

In die Oberflächenschicht eines schwach P-dotierten Siliziumkristalls („Body“) werden zwei stark N-dotierte Zonen („Source“ und „Drain“) eingebracht. Das Gebiet zwischen diesen beiden Zonen ist mit einer wenige hundert Atomlagen dünnen Isolierschicht aus Quarzglas (SiO_2) bedeckt, auf die eine metallisch leitende Steuerelektrode („Gate“) aus Aluminium aufgedampft wird, die mit den Source- und Drain-Zonen etwas überlappt.

Da in der Praxis vergleichsweise hohe Oxidladungen $Q_{ox} > Q_{ss} + Q_{si}$ auftreten, wäre diese Struktur an sich vom Verarmungstyp. Durch eine ausreichend negative Vorspannung des P-leitenden Body gegenüber der N-leitenden Source, wodurch die Breite der Sperrschicht zwischen beiden zunimmt, kann der spontane N-leitende Kanal beseitigt und als *Anreicherungstyp* betrieben werden. Aus technologischen Gründen ist dies der am häufigsten verwendete MOS-Transistortyp.

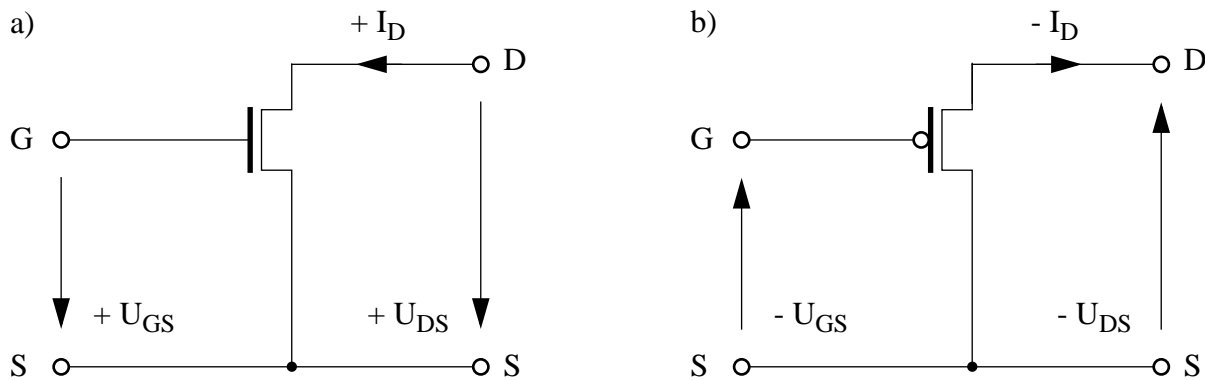


Bild 5.11: Schaltsymbole von MOS-Transistoren; G: Gate, S: Source, D: Drain.

a) N-Kanal, Anreicherungstyp, $U_{thN} > 0$ b) P-Kanal, Anreicherungstyp, $U_{thP} < 0$

Wird das Gate gegenüber der Source soweit positiv vorgespannt, dass die Schwellspannung überschritten wird ($U_{GS} > U_{th}$), so wird die Elektronendichte unter dem Gate und damit zwischen Source und Drain soweit erhöht, dass an der Oberfläche des P-leitenden Body der Leitungstyp invertiert wird und ein N-leitender Kanal entsteht, der Source mit Drain leitend verbindet. Bei Anlegen einer Spannung $U_{DS} > 0$ zwischen Drain und Source fließen Elektronen von der Source („Quelle“) zur Drain („Senke“) und damit ein definitionsgemäß positiver Drainstrom I_D (Bild 5.11).

In der Digitaltechnik wird der MOS-Transistor als Schalter mit zwei wohlunterschiedenen Zuständen („binär“) betrieben. Man schaltet dazu die Eingangsspannung U_{GS} zwischen zwei diskreten Werten, die beidseits der Schwellspannung U_{th} liegen. Der Kanal wechselt dann zwischen dem leitenden und dem sperrenden Zustand.

5.4.2 Kennlinien und Betriebsbereiche

Beim MOS-Transistor unterscheidet man nur zwei Kennlinien (da wegen der isolierten Gate-Elektrode eine Eingangskennlinie $I_G = f(U_{GS}) = 0$ trivial wäre):

- **Die Steuerkennlinie** $I_D = f(U_{GS})$ für $U_{DS} = \text{const.}$ Die Ansteuerspannung U_{GS} muss die Schwellspannung U_{th} überschreiten, damit ein meßbarer Drainstrom I_D fließt, der mit zunehmender Ansteuerspannung U_{GS} ansteigt. Voraussetzung für einen Drainstromfluß ist ferner eine Drain-Source-Spannung $U_{DS} > 0$.
- **Das Ausgangskennlinienfeld** $I_D = f(U_{DS})$ für $(U_{GS} - U_{th}) = \text{const.}$ als Parameter. Erhöht man bei konstanter Gate-Vorspannung U_{GS} die Drain-Source-Spannung U_{DS} , so steigt der Drainstrom I_D zunächst ebenfalls an. Mit zunehmendem Spannungsabfall U_{DS} längs des Kanals wird dieser am drainseitigen Ende immer dünner. Dadurch wird der Anstieg des Drainstroms I_D allmählich geringer. Bei weiter zunehmender Drainspannung U_{DS} wird der Kanal drainseitig schließlich abgeschnürt, wobei er sich immer enger an die Si-SiO₂-Grenzschicht anschmiegt. Steigt U_{DS} über diese Abschnürspannung U_p („pinch-off voltage“) an, fällt der zusätzliche Spannungsbetrag an der sich verlängernden Abschnürung des Kanals ab, so dass der Drainstrom I_D nicht mehr zunimmt, sondern einen Sättigungswert erreicht.

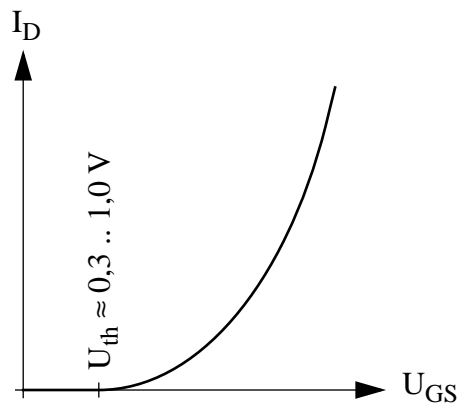


Bild 5.12: Steuerkennlinie eines MOS-Transistors mit N-Kanal

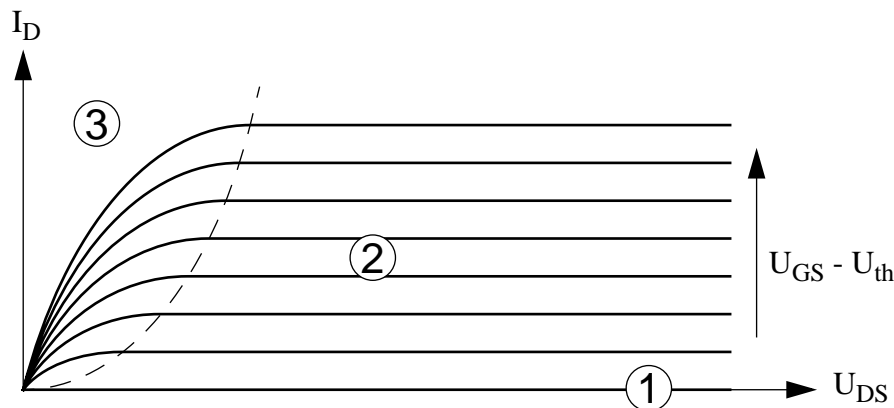


Bild 5.13: Ausgangskennlinienfeld eines MOS-Transistors mit N-Kanal;
1 - gesperrter Bereich, 2 - gesättigter Bereich, 3 - linearer Bereich^{*)}

Ein Vergleich der Ausgangskennlinienfelder des bipolaren Transistors (Bild 5.6) und des MOS-Transistors (Bild 5.13) läßt erkennen, dass beide näherungsweise als Stromquellen $I = \text{const.}$ wirken, obwohl die physikalischen Erklärungen völlig unterschiedlich sind. Beim MOS-Transistor kann man drei Betriebsbereiche unterscheiden:

- **Gesperrter Bereich:** Es gilt $U_{GS} < U_{th}$. Es besteht kein leitender Kanal zwischen Source und Drain („geöffneter Schalter“).
- **Gesättigter Bereich:** Es gilt $U_{GS} \geq U_{th}$, aber $U_{DS} > U_p$ (Abschnürspannung, Gl. 5.13). Hier gilt für den Drainstrom:

$$I_{D, \text{sat}} = \frac{\mu}{2} \cdot c_{ox} \cdot \frac{W}{L} \cdot (U_{GS} - U_{th})^2 \quad (5.11)$$

- **Linearer Bereich:** Es gilt $U_{GS} \geq U_{th}$, während $U_{DS} < U_p$ (Abschnürspannung, Gl. 5.13). Hier gilt für den Drainstrom („geschlossener Schalter“):

$$I_{D, \text{lin}} = \mu \cdot c_{ox} \cdot \frac{W}{L} \cdot \left(U_{GS} - U_{th} - \frac{1}{2} U_{DS} \right) \cdot U_{DS} \quad (5.12)$$

^{*)} Man beachte, dass die Bereiche anders benannt sind als für den bipolaren Transistor in Bild 5.6.

Dabei ist μ die Beweglichkeit der Ladungsträger im Kanal (hier Elektronen), c_{ox} die Kapazität pro Flächeneinheit im Kanalbereich. W ist die Breite des Kanalbereichs und L dessen Länge, d.h. der Drainstrom hängt von der Auslegung („Layout“) des MOS-Transistors ab.

Die Grenze zwischen linearem und gesättigtem Bereich bezeichnet die Abschnürspannung U_P . Durch Gleichsetzen von Gl.(5.11) und Gl.(5.12) erhält man:

$$I_{D,p} = \frac{\mu}{2} \cdot c_{ox} \cdot \frac{W}{L} \cdot U_P^2 \quad \text{wobei } U_P = U_{DS} = (U_{GS} - U_{th}) \quad (5.13)$$

Für die Schwellspannung gilt schließlich:

$$U_{th} = \frac{1}{c_{ox}} \cdot (K_1 \sqrt{\Phi - U_{BS}} - K_2) \quad c_{ox} = \epsilon_{ox} \cdot \epsilon_0 / d_{ox} \quad (5.14)$$

Wie man sieht ist die Schwellspannung U_{th} umgekehrt proportional zur Flächenkapazität c_{ox} im Kanalbereich. Diese wiederum hängt bekanntlich von der Dicke d_{ox} des Gate-Oxids (SiO_2) und von $\epsilon_{ox} = 3,9$ ab, der relativen Dielektrizitätskonstanten des SiO_2 . $\Phi = 0,5 \dots 1,0 \text{ V}$ ist eine Materialkenngröße; sie muss wie die Faktoren K_1 und K_2 im Einzelfall meßtechnisch ermittelt werden. Gl.(5.14) gibt die bereits erwähnte Empfindlichkeit der Schwellspannung $U_{th} = f(U_{BS})$ wieder, d. h. ihre Abhängigkeit von der Body-Vorspannung U_{BS} .

6.1 Ersatzschaltbilder der Mikroelektronik

Die Fertigungsverfahren der Mikroelektronik sind auf eine Massenfertigung integrierter Schaltungen abgestimmt. Im Gegensatz zur konventionellen Schaltungstechnik mit einzelnen, diskreten elektronischen Bauelementen ist es in der Mikroelektronik nicht möglich, eine neu entworfene Schaltung als Prototyp in wenigen Tagen zu realisieren, um ihr Verhalten überprüfen und die Schaltung gegebenenfalls ändern zu können. Vielmehr benötigt man zur Realisierung einige Wochen, erhält dann allerdings eine außerordentlich hohe Stückzahl gleichartiger Einheiten. Aus wirtschaftlichen Gründen muss deshalb das Verhalten einer neu entworfenen Schaltung durch Simulation soweit vorausberechnet und der Schaltungsentwurf unter Umständen vorab modifiziert werden, damit bereits die erste gefertigte Serie das gewünschte Verhalten zeigt. Das ist insbesondere bei integrierten Schaltkreisen, deren in einem gemeinsamen Halbleiterkristall untergebrachte Bauelemente zahlreiche Neben- und Wechselwirkungen aufweisen, praktisch nur unter Zuhilfenahme von Entwurfsprogrammen möglich („Computer Aided Design“, CAD). Ferner erlaubt nur der rechnergestützte Entwurf eine Berücksichtigung von Fertigungstoleranzen der elektrischen Parameter. Das ist wichtig, da es beim Entwurf in Massen zu fertigender Schaltkreise nicht darauf ankommt, dass ein Schaltkreis möglichst gut, sondern dass möglichst viele Schaltkreise ausreichend gut arbeiten. Die erwünschten sowie die unerwünschten („parasitären“) Eigenschaften der integrierten Bauelemente müssen daher in *Ersatzschaltbildern* erfaßt werden, die aus idealen, linearen, passiven elektrischen Schaltelementen bestehen. In der Digitaltechnik sind spezielle Ersatzschaltbilder von Interesse, die für den Betrieb mit Signalen hoher Amplitude gelten. Wegen des Umfangs und der Komplexität hochintegrierter digitaler Schaltkreise sind relativ einfache Ersatzschaltbilder zu wählen, um den Schaltungsentwurf in angemessener Zeit und mit vertretbarem Aufwand durch Simulation validieren zu können. Ersatzschaltbilder sollen vor allem die parasitären Eigenschaften der nachgebildeten Strukturen erfassen. Bei integrierten Schaltkreisen sind die Nebenwirkungen unter Umständen so ausgeprägt, dass zunächst störende Effekte bei geeignetem Entwurf der Schaltung sogar als Nutzeffekte verwendet werden können. Es werden daher in den folgenden Abschnitten die Ersatzschaltbilder integrierter Bauelemente vorgestellt und diskutiert. Die elektronischen Bauelemente integrierter Schaltkreise sind im wesentlichen aus folgenden Strukturen aufgebaut:

- Bahn aus dotiertem Halbleitermaterial oder Metall → Widerstand in Bipolartechnologie
- Einzelner PN-Übergang → Diode bzw. Kondensator in Bipolartechnologie
- Zwei eng benachbarte PN-Übergänge, einer gesperrt und einer leitend → bipolarer Transistor
- Halbleiterkanal, der zwei gesperrte PN-Übergänge leitend miteinander verbindet → MOS-Transistor und „Widerstand“ in MOS-Technologie
- Schichtfolge aus zwei elektrisch leitenden Lagen mit dazwischen liegender Isolierschicht → Kondensator in MOS-Technologie

In der beigegeführten Übersicht sind diese Strukturen zusammen mit den entsprechenden ersatzweisen Schaltsymbolen wiedergegeben. Daraus lassen sich sämtliche Ersatzschaltbilder integrierter Bauelemente je nach ihrem Aufbau zusammensetzen. Die Werte der einzelnen Ersatzelemente sind allerdings in den wenigsten Fällen konstant; sie hängen unter Umständen stark vom Betriebszustand des betreffenden Bauelements ab. Es ist jedoch möglich, diese Abhängigkeit beim rechnergestützten Schaltungsentwurf nachzubilden.

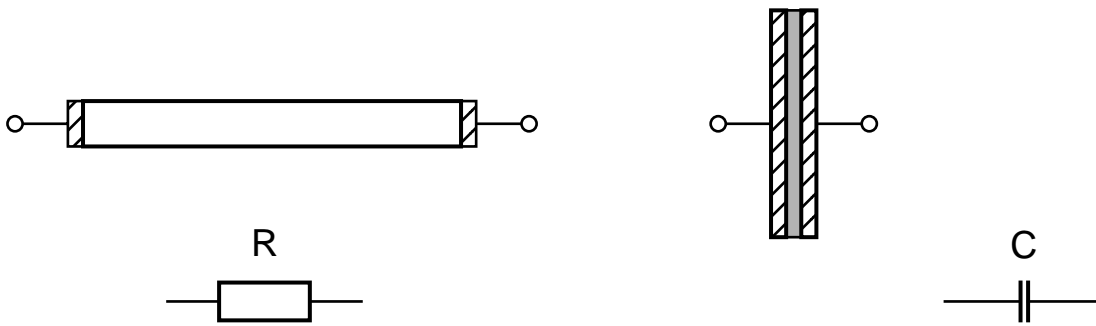


Bild 6.1: Widerstandsbahn und Schaltsymbol

Kondensatorstruktur und Schaltsymbol

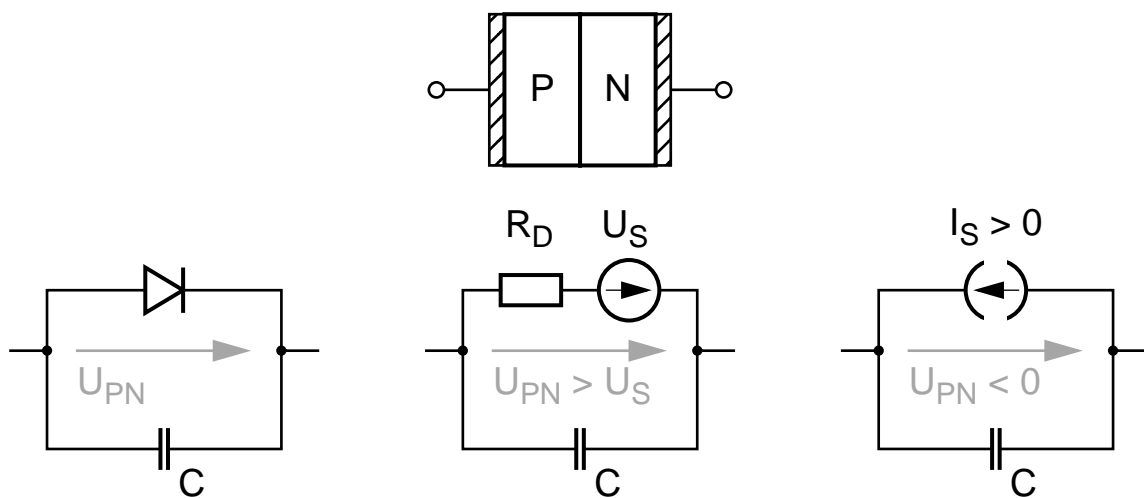


Bild 6.2: PN-Übergang; allgemeine Diode (links), leitende Diode (mitte), gesperrte Diode (rechts)

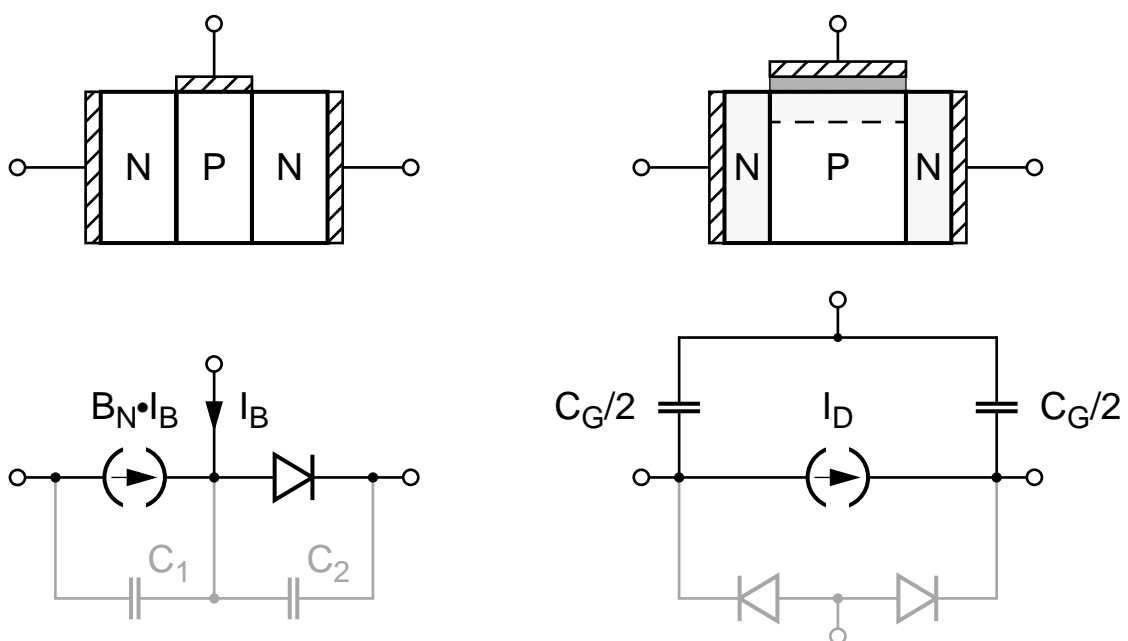


Bild 6.3: Bipolarer Transistor

MOS-Transistor

6.2 Bauelemente in Bipolartechnologie

Bei der Herstellung von Mikroschaltkreisen werden Millionen elektronischer Bauelementen - wie Transistoren, Widerstände, Kondensatoren und Dioden - zu untrennbaren Schaltungseinheiten zusammengefasst („Höchstintegration“), die wiederum als Bausteine für größere Funktionseinheiten dienen. Bei dieser Technologie werden alle aktiven und passiven Bauelemente in einem gemeinsamen Halbleiterkristall erzeugt.

- **Aktive Bauelemente.** Zur Ausführung digitaler Schaltvorgänge benötigt man Bauelemente mit mindestens drei Anschlüssen. Zwischen zwei Anschlüssen befindet sich ein elektrischer Pfad, der durch den dritten Anschluss zwischen einem hoch- und einem niederohmigen Zustand schaltbar ist. Man verwendet in der Mikroelektronik hierzu entweder bipolare oder MOS-Transistoren. Die Fertigungsabläufe werden auf die aktiven Bauelemente abgestimmt.
- **Passive Bauelemente.** Nicht schaltbare Bauelemente besitzen nur zwei Anschlüsse. In der Mikroelektronik handelt es sich vor allem um Widerstände, Kondensatoren und Dioden; Induktivitäten können in integrierter Technologie nicht hergestellt werden. Zur Herstellung passiver Bauelemente führt man aus wirtschaftlichen Gründen keine zusätzlichen Fertigungsschritte ein, so dass Transistoren mit entsprechender Verdrahtung oder auch Teile davon als passive Zweipole verwendet werden müssen.

6.2.1 Bipolare Transistoren vom NPN-Typ

a) Aufbau bipolarer Transistoren

Die wichtigsten Bauelemente einer Digitalschaltung sind die Transistoren. Sie werden zur Ausführung von Schaltvorgängen, zur Impulsauffrischung und zur Verstärkung von Signalen benötigt.

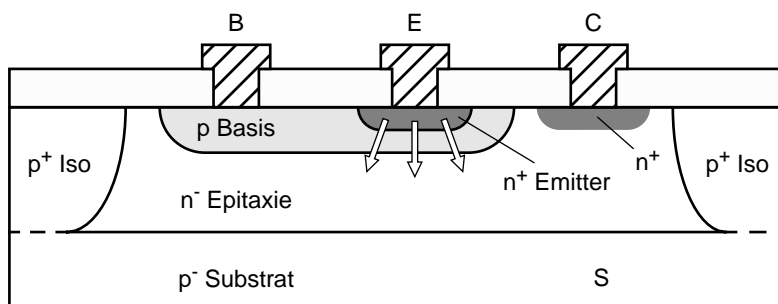


Bild 6.4: Querschnitt durch eine bipolare Transistorstruktur

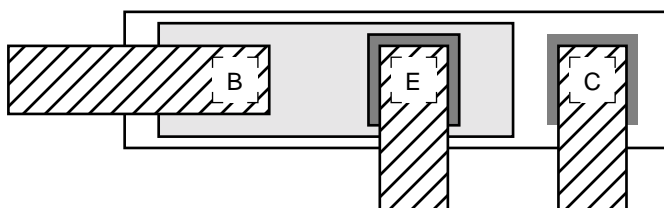


Bild 6.5: Draufsicht („Layout“) einer bipolaren Transistorstruktur

In der Bipolartechnologie werden die vertikalen Abmessungen durch das Fertigungsverfahren festgelegt, wie der vorhergehende Querschnitt zeigt. Beim Schaltungsentwurf erfolgt dann die horizontal-geometrische Auslegung („Layout“) der aktiven Bauelemente wie in der obigen Draufsicht.

Aus der Fülle charakteristischer Transistoreigenschaften sollen hier nur diejenigen herausgegriffen werden, die durch die geometrische Auslegung der Bauelemente beeinflusst werden können. Ein idealer bipolarer Transistor kann durch seine *Stromverstärkung* vollständig beschrieben werden. In der Digitaltechnik, d.h. im Impulsbetrieb, interessieren vor allem die Gleichstrom-Verstärkungsfaktoren. Dafür stehen zwei völlig gleichwertige Definitionen zur Verfügung:

$$A(\text{lpha}) = I_C / I_E \qquad B(\text{eta}) = I_C / I_B \qquad (6.1)$$

Sie lassen sich unter Berücksichtigung der Knotenregel für den Collectorstrom I_C , den Basisstrom I_B und den Emitterstrom I_E zusammenfassen:

$$B = \frac{A}{1 - A} \qquad (6.2)$$

Im Idealfall sind die Verstärkungsfaktoren im gesamten Betriebsbereich konstant. Bei realen bipolaren Transistoren hängen sie jedoch nicht zuletzt von der geometrischen Auslegung ab. Betrachten wir dazu den Querschnitt durch eine bipolare Transistorstruktur in Bild 6.4. Der eingezeichnete Pfeil soll die Injektion von Elektronen durch den Emitter in die Basis und die Überschwemmung des gesperrten Basis-Collector-Übergangs andeuten. Diesen Betriebszustand bezeichnet man als *Normalbetrieb*. Man erkennt, dass die meisten injizierten Ladungsträger vom Collector aufgenommen werden können, woraus sich nach Gl.(6.1) eine Stromverstärkung im Normalbetrieb A_N ergibt, die nahe bei Eins liegt; mit Gl.(6.2) ergeben sich dann Werte für B_N in folgender Größenordnung:

$$A_N \approx 1 \leq 1 \qquad B_N = 10 \dots 100$$

b) Ersatzschaltbilder bipolarer Transistoren

Als Beispiel soll hier der allgemein gebräuchliche NPN-Transistor betrachtet werden. Im Normalbetrieb des Transistors im *linearen* Bereich seines Ausgangskennlinienfeldes $I_C = f(U_{CE}; U_{BE})$ ist der Basis-Emitter-Übergang in Durchlaßrichtung, der Basis-Collector-Übergang in Sperrrichtung gepolt. Man betrachte Bild 6.6:

- Der *gesperrte* Basis-Collector-Übergang kann durch die Parallelschaltung seiner *Sperrschichtkapazität* C_{BC} und seines Sperrstroms $-I_{BC}$ ersatzweise dargestellt werden,
- der *leitende* Basis-Emitter-Übergang durch eine *Diffusionskapazität* C_{BE} und eine leitende Diode, deren nichtlineare Durchlaßkennlinie durch den Widerstand R_{BE} im jeweiligen Arbeitspunkt und die physikalisch gegebene Schleusenspannung $U_S \approx 0,7$ Volt der Siliziumdiode nachgebildet wird.
- Die Überschwemmung des gesperrten Basis-Collector-Übergangs mit Elektronen, die vom Emitter ins Basisgebiet injiziert werden - der eigentliche *Transistoreffekt* - wird durch die Stromquelle $B_N \cdot I_B$ erfaßt.

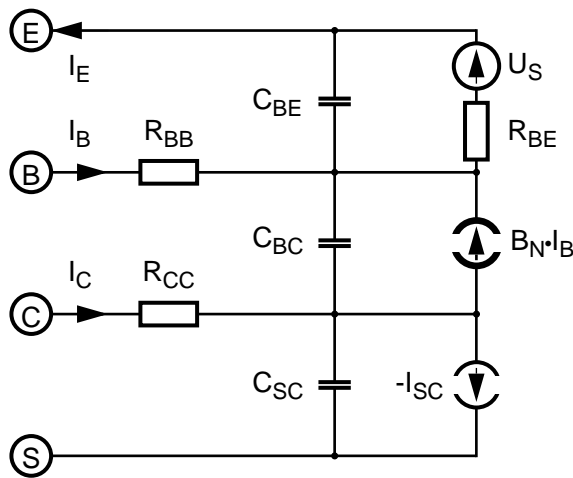


Bild 6.6: Ersatzschaltbild eines NPN-Transistors bei normalem Betrieb im linearen Bereich

- Die Bahnwiderstände zwischen den Anschlusskontakten und den PN-Übergängen werden durch die Widerstände R_{BB} und R_{CC} berücksichtigt; der Bahnwiderstand des Emittergebiets kann vernachlässigt werden.
- Die „Isolationswanne“, die von gesperrten PN-Übergängen gebildet wird, kann durch die Parallelschaltung einer Sperrschichtkapazität C_{SC} und eines Sperrstromes $-I_{SC}$ dargestellt werden. Der Kapazitätswert von C_{SC} hängt von der Struktur der Isolationswanne ab. Nach Bild 6.4 wird sie von zwei verschiedenen PN-Übergängen gebildet: zwischen der Epitaxieschicht einerseits und dem Substrat sowie der Isolationswand (p^+ Iso) andererseits, entsprechend kompliziert ist ihr elektrisches Verhalten.

Wird ein bipolarer Transistor *invers* betrieben, so ist der normale Basis-Collector-Übergang in Durchlaßrichtung, der normale Basis-Emitter-Übergang in Sperrichtung gepolt. Die Überschwemmung des nun gesperrten Basis-Emitter-Übergang mit Elektronen ist dann durch eine Stromquelle $B_I \cdot I_B$ zu erfassen. Zusätzlich kann ein parasitärer „Substrattransistor“ in Aktion treten. Invers zum normalen Ersatzschaltbild nach Bild 6.6 kann eines für den inversen Betrieb erstellt werden.

Wird der Transistors im *gesättigten* Bereich seines Ausgangskennlinienfeldes $I_C = f(U_{CE}; U_{BE})$ betrieben, so sind sowohl der Basis-Emitter-Übergang als auch der Basis-Collector-Übergang in Durchlaßrichtung gepolt. Beide N-leitenden Gebiete injizieren Elektronen ins Basisgebiet aus entgegengesetzten Richtungen. Auch dabei kann ein parasitärer Substrattransistor entstehen. Das gesamte Ersatzschaltbild setzt sich dann aus den Ersatzschaltbildern für den Normal- und den Inversbetrieb zusammen. Dabei können Sperrschichtkapazitäten und Sperrströme gegenüber Diffusionskapazitäten und gesteuerten Stromquellen vernachlässigt werden.

6.2.2 Passive Bauelemente in Bipolartechnologie

a) Aufbau integrierter Widerstände

Widerstände werden in der digitalen Schaltungstechnik vor allem als Lastelemente in Inverter-, Verknüpfungs- und Speicherschaltungen, für Spannungsteiler und als Koppellemente verwendet.

In der Bipolartechnologie erzeugt man ohmsche Widerstände, indem man mit demselben Fertigungsschritt, bei dem auch die Basisgebiete der Transistoren entstehen, Bahnen der erforderlichen Länge und Breite in eine Isolationswanne, d.h. die Epitaxieschicht einbringt. Bild 6.7 zeigt einen Querschnitt durch einen integrierten Widerstand. Er muss so betrieben werden, dass der PN-Übergang stets gesperrt ist, den er zur Epitaxieschicht bildet. Falls die Potentiale schaltungstechnisch so gewählt werden, können zur Platzersparnis mehrere basisdotierte Widerstände in einer gemeinsamen Isolationswanne untergebracht werden.

Die elektrische Leitfähigkeit σ des Halbleitermaterials ist bekanntlich proportional zur Dichte n und Beweglichkeit μ der Ladungsträger, Proportionalitätsfaktor ist die Elementarladung e :

$$\sigma = e n \mu \qquad \sigma = e |n_A - n_D| \mu \qquad (6.3)$$

Im Temperaturbereich der Störstellenerschöpfung ist die Ladungsträgerdichte gleich der Netto-Störstellendichte: $n = |n_A - n_D|$. Durch entsprechendes Dotieren lässt sich somit der spezifische Widerstand $\rho = 1 / \sigma$ des Halbleitermaterials einstellen.

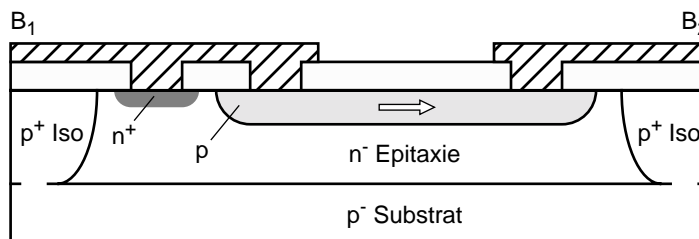


Bild 6.7: Querschnitt eines integrierten Widerstandes

Bei der geometrischen Konstruktion einer integrierten Schaltung („Layout“) verwendet man als charakteristische Größe den Schichtwiderstand R_S , d.h. den auf die Schichtdicke d bezogenen spezifischen Widerstand ρ des Widerstandsmaterials:

$$R_S = \frac{\rho}{d} \qquad R = R_S \cdot \frac{L}{W} \qquad (6.4)$$

Beide Werte für ρ und d sind durch die Technologie festgelegt und nur die Länge L und Breite W der Widerstandsbahn können beim Schaltungsentwurf gewählt werden (Bild 6.8).

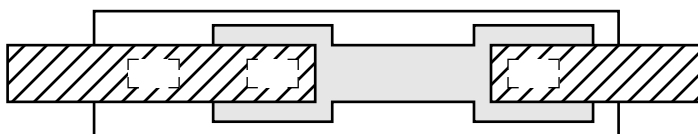


Bild 6.8: Draufsicht („Layout“) eines integrierten Widerstandes

Der Schichtwiderstand R_S entspricht dem Widerstandswert einer quadratischen Auslegung; er wird daher in Ω/\square angegeben. Da die entsprechenden Störstellendichten und Eindringtiefen auf die Transistorstruktur optimiert werden, sind die Schichtwiderstände von vornherein auf einen vorgegebenen Wertebereich eingeschränkt:

$$R_S (\text{Basis}) \approx 100 \dots 400 \Omega/\square$$

b) Ersatzschaltbilder integrierter Widerstände

Ein basisdotierter Widerstand bildet zur Epitaxieschicht, in die er eingebettet ist, einen im Betrieb gesperrten PN-Übergang. Er ist daher mit einer Sperrschichtkapazität C_{BC} und einem Sperrstrom $-I_{BC}$ behaftet, die entlang der Widerstandsbahn R verteilt sind. Wird wie in Bild 6.9 die N-leitende Epitaxieschicht (C) mit (B_1), dem positiveren der beiden Widerstandsanschlüsse kurzgeschlossen, so ist die Hälfte der Sperrschichtkapazität C_{BC} bzw. des Sperrstroms $-I_{BC}$ unwirksam. Zwischen der Epitaxieschicht (C) und dem Substrat (S) befinden sich noch die Sperrschichtkapazität C_{SC} und der Sperrstrom $-I_{SC}$ der Isolationswanne. Das P-leitende Substrat (S) liegt stets auf dem negativsten in der Schaltung vorkommenden Potential.

Würde die Epitaxieschicht nicht angeschlossen, so stellte sich ihr Potential frei ein, der PN-Übergang zwischen Widerstandsbahn (B_2) und Epitaxieschicht (C) wäre dann schwach in Durchlaßrichtung gepolt. Die Schichtfolge aus P-Widerstandsbahn, N-Epitaxieschicht und P-Substrat wirkte dann als parasitärer „Substrattransistor“ und verschlechterte die elektrischen Eigenschaften des integrierten Widerstandes ganz erheblich.

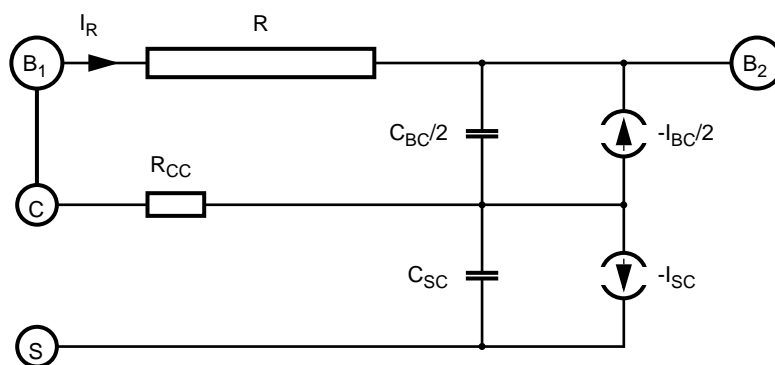


Bild 6.9: Ersatzschaltbild eines integrierten Widerstandes

c) Aufbau integrierter Kondensatoren

Im Prinzip besteht ein Kondensator aus zwei eng benachbarten, elektrisch gut leitenden Gebieten, die durch eine dünne nichtleitende Schicht voneinander isoliert sind. In der Bipolartechnologie bieten sich hierfür gesperrte PN-Übergänge an. Es fließt dann allerdings ein kleiner Sperrstrom, der den Kondensator laufend entlädt und so seine Güte begrenzt. Kondensatoren können auf Grund ihrer Fähigkeit, Energie zu speichern, in der digitalen Schaltungstechnik zur Informationsspeicherung, zum Aufbau dynamischer Verknüpfungsschaltungen, als zeitbestimmende Glieder sowie ganz allgemein zur Kopplung oder Entkopplung verwendet werden.

Eine charakteristische Größe beim Entwurf integrierter Schaltungen ist die Flächenkapazität C_F . Den Kapazitätswert eines Kondensators erhält man durch Multiplizieren der Fläche des betreffenden PN-Übergangs mit der Flächenkapazität:

$$C_F = \frac{\epsilon}{d} \qquad C = C_F \cdot W \cdot L \qquad (6.5)$$

Dielektrizitätskonstante ϵ und Dicke d der „Isolierschicht“ hängen von den beteiligten Störstellendichten ab, also von der Technologie, und von der angelegten Sperrspannung, bei deren Erhöhung sich die Sperrschicht bekanntlich verbreitert, so dass die wirksame Dicke d entsprechend zunimmt. Es handelt sich somit um spannungsabhängige Kapazitäten. Da die Dotierungen in erster Linie auf die bipolare Transistorstruktur optimiert wird, sind die Werte für die Flächenkapazitäten von vornherein vorgegeben.

In Bipolartechnologie stehen zur Auslegung von Kondensatoren zwei unterschiedliche PN-Übergänge zur Verfügung, und zwar die des Transistors: der Basis-Emitter- und der Basis-Collector-Übergang. Integrierte Kondensatoren erhält man durch Verdrahten der bipolaren Transistorstruktur zu einem Zweipol; Bild 6.10 zeigt ein Beispiel. Beide PN-Übergänge werden parallel geschaltet, um bei gleichem Platzbedarf eine höhere Gesamtkapazität zu erreichen.

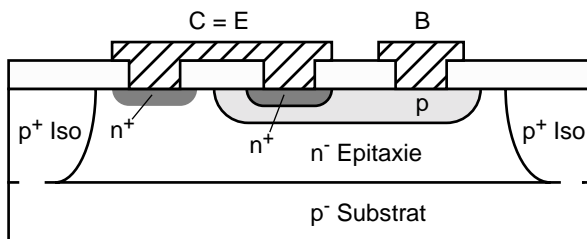


Bild 6.10: Querschnitt eines integrierten Kondensators mit gesperrten PN-Übergängen

d) Aufbau integrierter Dioden

Die nichtlineare Strom-Spannungs-Kennlinie einer Diode legt vor allem folgende Anwendungen in der digitalen Schaltungstechnik nahe:

- Bei Stromeinprägung im Durchlaßbereich kann die Diode zur Potentialverschiebung um den Wert ihrer Schleusenspannung $U_S \approx 0,7$ Volt dienen.
- Der kleine Sperrstrom kann zur Kompensation der Leckströme von Speicherschaltungen verwendet werden.
- Als Koppелеlemente in Verknüpfungsschaltungen und als Schreib/Lese-Elemente in Speicherschaltungen werden Dioden laufend zwischen Durchlaß- und Sperrbereich umgeschaltet.

In Bipolartechnologie stehen zur Herstellung von Dioden wieder die PN-Übergänge des Transistors zur Verfügung: der Basis-Emitter- und der Basis-Collector-Übergang, die unterschiedliche Strom-Spannungs-Kennlinien aufweisen. Integrierte Dioden erhält man durch Verdrahten der bipolaren Transistorstruktur zu einem Zweipol; Bild 6.11 zeigt ein Beispiel. Der nicht benötigte Collectoranschluss (C) wird mit dem Basisanschluss (B) verbunden, was über eine Transistorwirkung Einfluß auf die Durchlaßkennlinie der integrierten Diode hat.

Da jede integrierte Diode zwangsläufig mit einer Sperrschichtkapazität behaftet ist, unterscheidet sich ihre Struktur nur schaltungstechnisch von derjenigen integrierter Kondensatoren. Während jedoch die PN-Übergänge bei Verwendung als Kondensator stets in Sperrichtung betrieben und so großflächig ausgelegt werden, wie es der benötigte Kapazitätswert erfordert, werden sie bei Verwendung als Diode platzsparend so klein ausgelegt, wie es die Strombelastbarkeit zuläßt, und auch in Durchlaßrichtung betrieben.

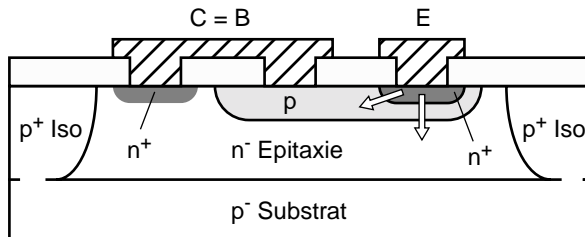


Bild 6.11: Querschnitt einer integrierten Diode mit leitendem Basis-Emitter-Übergang (man beachte den Transistoreffekt)

e) Ersatzschaltbilder integrierter Kondensatoren und Dioden

Da ein PN-Übergang grundsätzlich mit einer Kapazität behaftet ist, handelt es sich bei integrierten Kondensatoren und Dioden prinzipiell um dieselben Anordnungen, die sich nur in der ausgelegten Fläche unterscheiden. Somit sind die Ersatzschaltbilder im Sperrbereich identisch (Bild 6.12). Integrierte Kondensatoren sind in einer Isolationswanne untergebracht, so dass sie zwangsläufig mit einer relativ hohen Sperrschichtkapazität C_{SC} behaftet sind. Der Wert dieser Sperrschichtkapazität kann aus der Struktur der Isolationswanne berechnet werden, die nach Bild 6.10 durch zwei unterschiedliche PN-Übergänge gebildet wird: zwischen der Epitaxieschicht einerseits und dem Substrat sowie der Isolationswand (p^+ Iso) andererseits.

Im Fall einer integrierten Diode ist die Epitaxieschicht (C) mit dem Basisgebiet (B) kurzgeschlossen, um ihr elektrisches Potential festzulegen. Dadurch wird der Substrattransistor unwirksam. Bei dieser Schaltung wird aber im Durchlaßbereich der Diode der normale Transistor aktiv, die im Ersatzschaltbild durch eine Stromquelle $A_N \cdot I_E$ berücksichtigt werden muss (Bild 6.13). Bei Siliziumdioden ist charakteristisch, dass erst bei Anlegen einer Durchlaßspannung von mehr als der Schleusenspannung $U_S \approx 0,7$ Volt ein merklicher Durchlaßstrom fließt, der dann aber steil ansteigt.

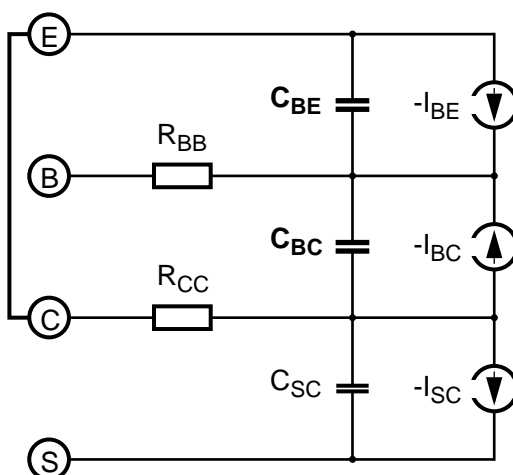


Bild 6.12: Ersatzschaltbild einer Diode im Sperrbereich oder eines integrierten Kondensators

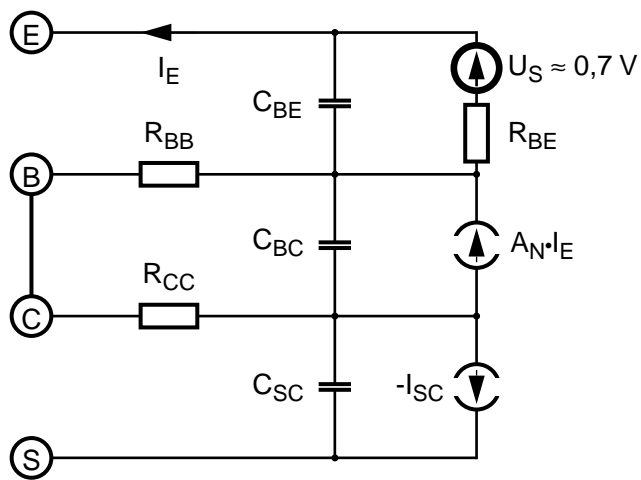


Bild 6.13: Ersatzschaltbild einer integrierten Diode im Durchlaßbereich

6.3 Bauelemente in MOS-Technologie

6.3.1 MOS-Transistoren vom N-Kanaltyp

a) Aufbau von MOS-Transistoren

In MOS-Technologie werden an der Oberfläche des Halbleiterkristalls horizontale, symmetrische Transistorstrukturen erzeugt (Bild 6.14), während in bipolarer Technologie vertikal angeordnete, unsymmetrische Transistoren im Innern des Halbleiterkristalls entstehen (vgl. Bild 6.4).

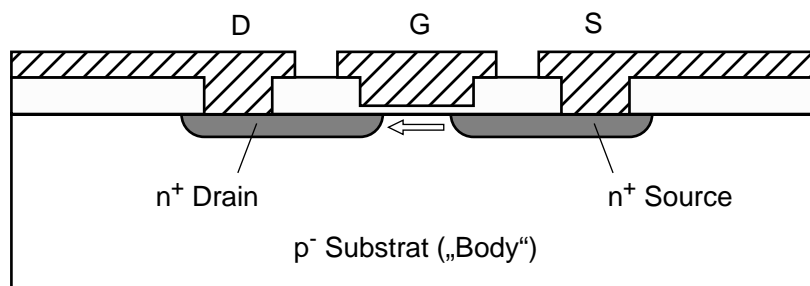


Bild 6.14: Querschnitt durch eine MOS-Transistorstruktur

Charakteristisch für das Verhalten eines MOS-Transistors ist zum einen seine Schwellspannung U_{th} („threshold voltage“), zum anderen die normierte Steilheit β :

$$U_{th} = \frac{1}{c_{ox}} (K_1 \sqrt{\Phi - U_{BS}} - K_2) \quad \beta = \mu c_{ox} \frac{W}{L} \quad (6.6)$$

Man erkennt aus Gl.(6.6), dass beide Kennwerte von der Flächenkapazität $c_{ox} = \epsilon / d$ im Gatebereich abhängen, wobei ϵ die Dielektrizitätskonstante und d die Dicke der Oxidschicht über dem Kanal ist. Der Fertigungsprozeß wird auf die im Gatebereich erforderliche Flächenkapazität c_{ox} abgestimmt. Beim Schaltungsentwurf erfolgt die horizontal-geometrische Auslegung („Layout“) der aktiven Bauelemente (Bild 6.15). Aus Gl.(6.6) entnimmt man ferner, dass die Steilheit β des MOS-Transistors durch die Auslegung von Weite W und Länge L des Kanalbereichs innerhalb weiter Grenzen frei gewählt werden kann. Es lassen sich somit hochohmige Strukturen mit langem, schmalem Kanal $L \gg W$ oder niederohmige mit breitem, kurzem Kanal $L \ll W$ auslegen.

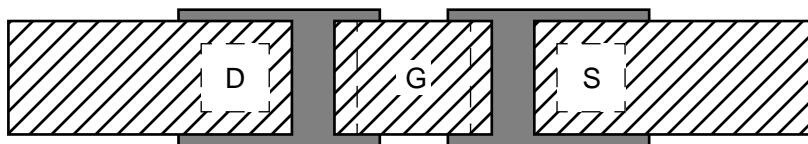


Bild 6.15: Draufsicht („Layout“) einer MOS-Transistorstruktur

b) Ersatzschaltbilder von MOS-Transistoren

Wichtigster Bestandteil des Ersatzschaltbildes eines MOS-Transistors mit isolierter Steuerelektrode ist der MOS-Kondensator, auf dessen Prinzip seine Wirkungsweise beruht. Dabei ist zwischen dem

Betriebszustand mit Kanalbildung zwischen Source und Drain und dem Zustand ohne Kanal zu unterscheiden. Als Beispiel soll hier der selbstsperrende N-Kanal-Typ betrachtet werden.

In folgenden Bild ist links das Ersatzschaltbild des *gesperrten* MOS-Transistors dargestellt.

- Die vom Substrat (Body) durch eine dünne Oxidschicht isolierte Steuerelektrode (Gate) bildet den genannten MOS-Kondensator C_{GB} . Sein Kapazitätswert ist durch die Flächenkapazität c_{ox} und die Fläche des dünnen Oxids über dem Kanalbereich gegeben. Die Technologie ist in der Regel vorgegeben, die Flächenkapazität c_{ox} im Gatebereich damit festgelegt.

$$C_{GB} = c_{ox} \cdot W \cdot L \qquad c_{ox} = \frac{\epsilon}{d} \qquad (6.7)$$

- Die Kapazitäten C_{GD} und C_{GS} sind die Überlappkapazitäten zwischen der Steuerelektrode (Gate) und den Drain- bzw. Sourcegebieten.
- Die gesperrten PN-Übergänge der Drain- und Sourcegebiete zum Substrat („Body“) sind durch die Sperrschichtkapazitäten C_{BD} und C_{BS} sowie durch die Sperrströme $-I_{BD}$ und $-I_{BS}$ nachgebildet. Die Erfassung der Sperrströme ist besonders wichtig, da sie bei gesperrtem MOS-Transistor das Potential einer freien Elektrode durch Entladung absenken können.

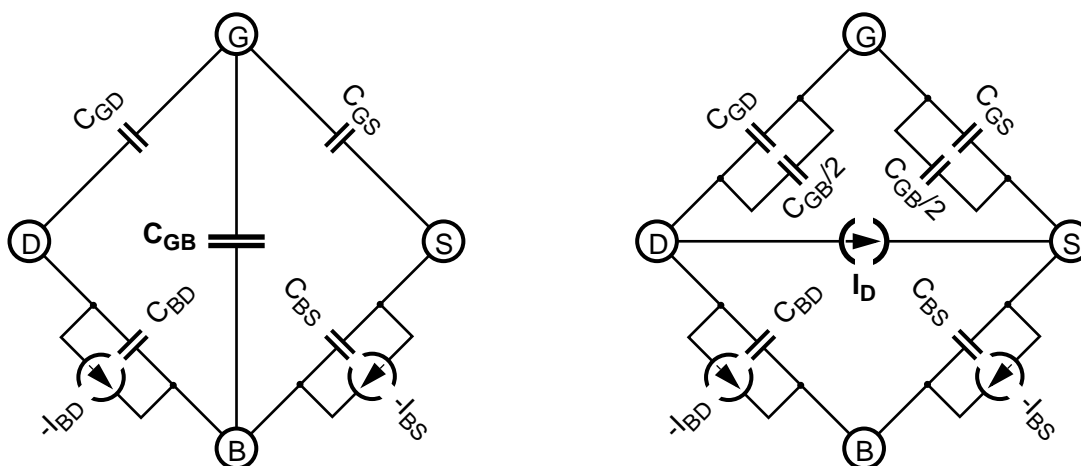


Bild 6.16: Ersatzschaltbilder eines MOS-Transistors; links: gesperrt, rechts: leitend

Besteht zwischen Drain (D) und Source (S) ein *leitender Kanal*, so kann er durch eine gesteuerte Stromquelle I_D im Ersatzschaltbild nachgebildet werden, wie im obigen Bild rechts gezeigt.

- Der Wert der Stromquelle I_D hängt vom Betriebszustand des MOS-Transistors ab, wobei besonders zu beachten ist, ob er im linearen oder im gesättigten Bereich arbeitet.

Für das Verhalten eines MOS-Transistors können folgende Strom-Spannungs-Gleichungen angegeben werden:

im gesättigten Bereich

$$I_{D, \text{sat}} = \frac{\beta}{2} (U_{GS} - U_{th})^2$$

im linearen Bereich

$$I_{D, \text{lin}} = \beta \left(U_{GS} - U_{th} - \frac{U_{DS}}{2} \right) U_{DS} \qquad (6.8)$$

- Durch den Kanal, der Drain und Source leitend verbindet, wird die Steuerelektrode (Gate) vom Substrat (Body) elektrisch abgeschirmt. Die MOS-Kapazität C_{GB} des Kanalbereichs liegt dann nicht mehr zwischen Gate (G) und Body (B), sondern zwischen Gate (G) und Kanal; sie kann im einfachsten Fall je zur Hälfte auf Drain (D) und Source (S) aufgeteilt werden.
- Die Sperrschichtkapazitäten und die Sperrströme, die entlang des leitenden Kanals zum Substrat (Body) hin verteilt sind, können den Sperrschichtkapazitäten C_{BD} und C_{BS} bzw. den Sperrströmen $-I_{BD}$ und $-I_{BS}$ zugerechnet werden.

6.3.2 Passive Bauelemente in MOS-Technologie

a) MOS-Transistoren als Widerstände

In MOS-Technologie steht als einzige die Source-Drain-Dotierung zur Verfügung, die einige $10 \Omega/\square$ Schichtwiderstand bietet. Es lassen sich damit nur niederohmige Widerstände zur Unterföhrung von Zwischenverbindungen bei der Verdrahtung einer integrierten Schaltung herstellen. Für mittlere und hochohmige Widerstände muss, wie im nächsten Bild, eine MOS-Transistorstruktur zum Zweipol verdrahtet werden, wobei Gate und Drain elektrisch leitend miteinander verbunden sind.

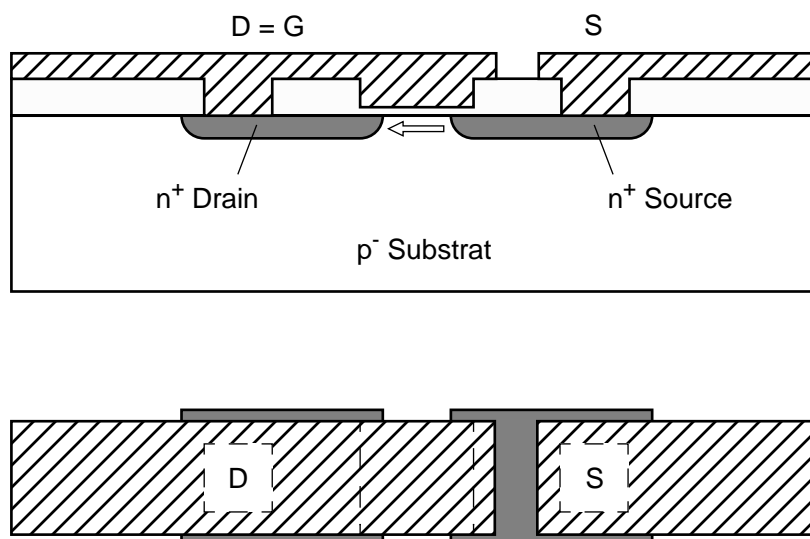


Bild 6.17: Zum Zweipol verdrahteter MOS-Transistor als „Widerstand“
oben: Querschnitt, unten: Draufsicht („Layout“)

b) Ersatzschaltbilder von MOS-Widerständen

Sind Gate (G) und Drain (D) kurzgeschlossen, dann gilt:

$$U_{DS} = U_{GS} > (U_{GS} - U_{th}) = U_P \quad I_{D, \text{sat}} = \frac{\beta}{2} U_P^2 \quad (6.9)$$

Der so zum Zweipol verdrahtete MOS-Transistor arbeitet im gesättigten Bereich nach Gl.(6.x), woraus die obige Gl.(6.x) folgt. Die Kennlinie und das zugehörige Ersatzschaltbild zeigt Bild 6.18.

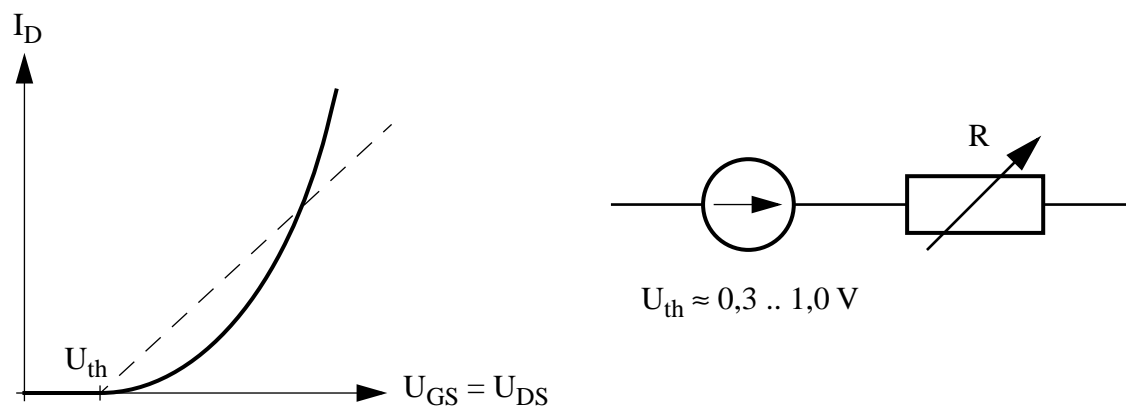


Bild 6.18: Kennlinie und Ersatzschaltung eines zum Zweipol verdrahteten MOS-Transistors

7.1 Inverterschaltungen

Die technische Informationsverarbeitung beruht auf der logischen Verknüpfung und Speicherung zweiwertig („binär“) codierter Informationen. Um sie mit elektrischen Signalen darzustellen, benötigt man zwei wohlunterschiedene („diskrete“) und somit abzählbare („digitale“) Signalpegel:

Lo \leftrightarrow niedriges Potential

Hi \leftrightarrow hohes Potential

Den beiden Signalpegeln kann man binäre Variablenwerte $\{0,1\}$ auf zweierlei Weise zuordnen:

positive Logik

Lo \leftrightarrow 0 ; Hi \leftrightarrow 1

negative Logik

Lo \leftrightarrow 1 ; Hi \leftrightarrow 0

7.1.1 Kontaktdarstellung von Transistoren

Bereits aus der Relaischnik stammt die Unterscheidung von zwei Kontakttypen:

- Arbeitskontakt - im Ruhezustand geöffnet, im Arbeitszustand geschlossen (Bild 7.1 links)
- Ruhekontakt - im Ruhezustand geschlossen, im Arbeitszustand geöffnet (Bild 7.2 links)

Ein Schalter besitzt per definitionem genau zwei Stellungen:

- geöffnet = sperrend, hochohmig, niedriger Stromfluß, hoher Spannungsabfall
- geschlossen = leitend, niederohmig, hoher Stromfluß, niedriger Spannungsabfall

Man vergleiche dazu die Kennlinien der beiden Kontakttypen in Bild 7.1 und Bild 7.2 rechts.

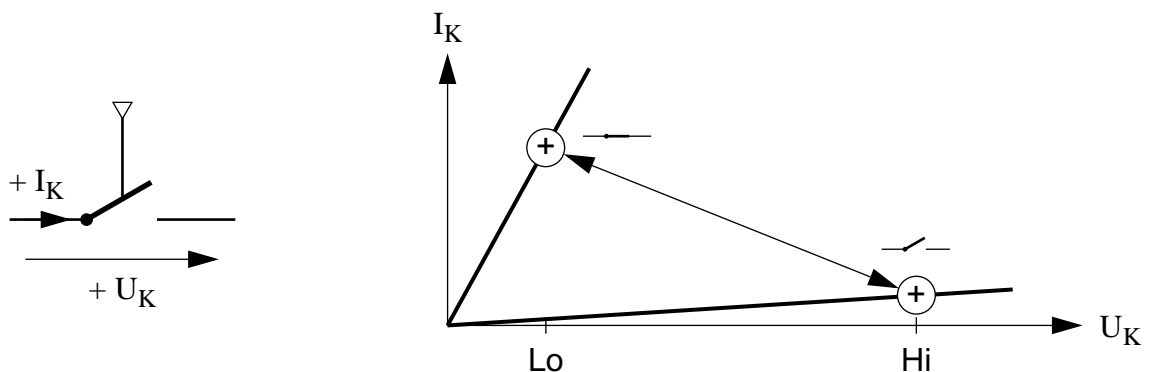


Bild 7.1: Arbeitskontakt mit Kennlinien

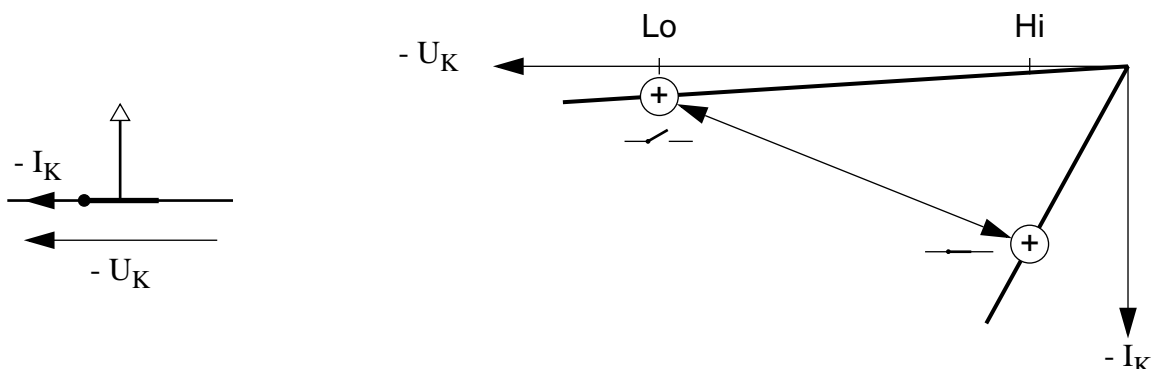


Bild 7.2: Ruhekontakt mit Kennlinien

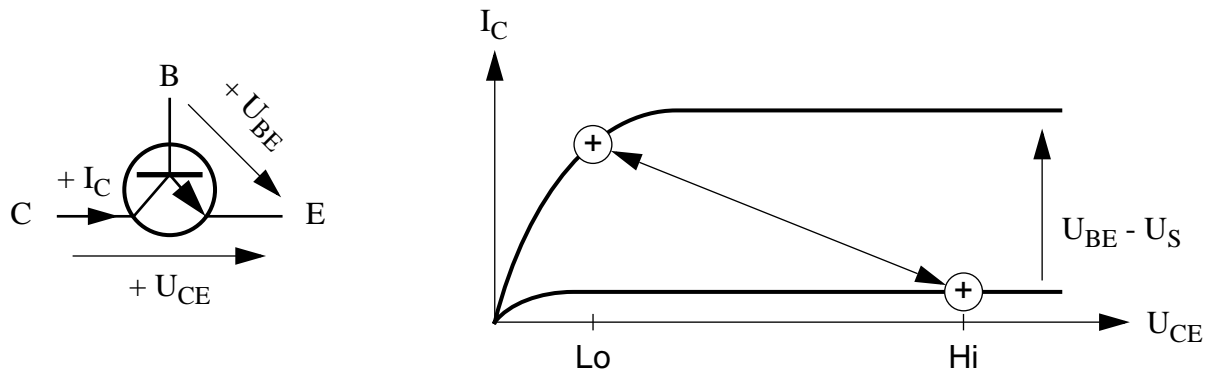


Bild 7.3: NPN-Transistor mit Ausgangskennlinien

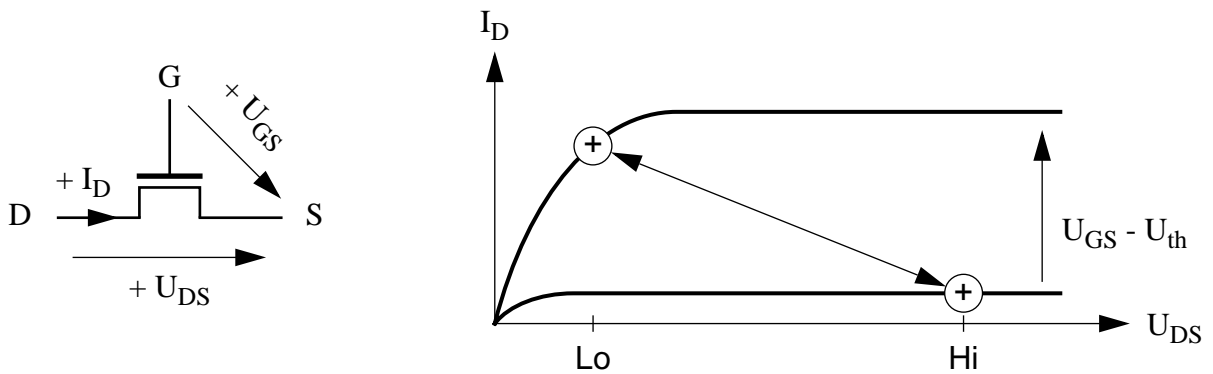


Bild 7.4: N-Kanal MOS-Transistor mit Ausgangskennlinien

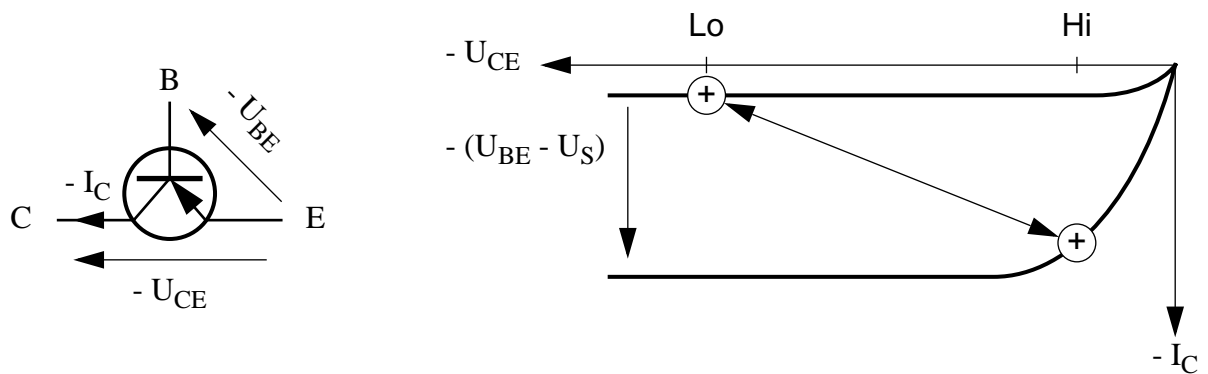


Bild 7.5: PNP-Transistor mit Ausgangskennlinien

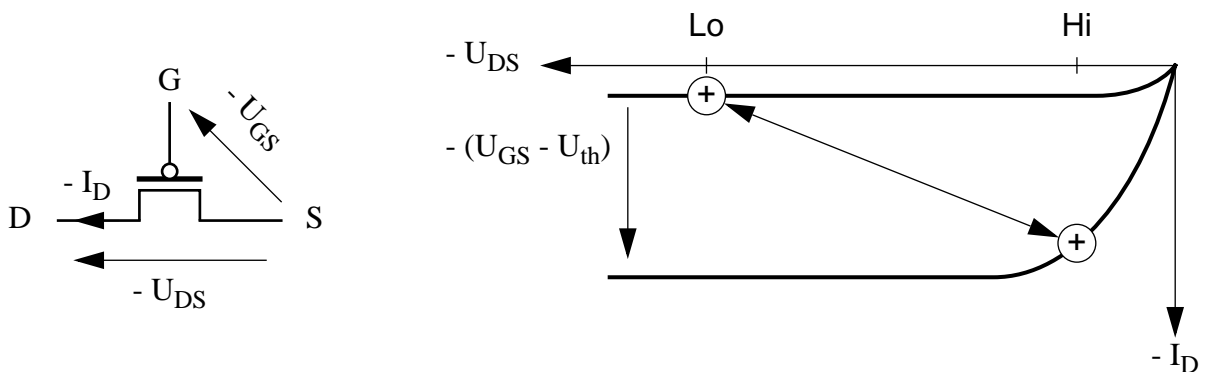


Bild 7.6: P-Kanal MOS-Transistor mit Ausgangskennlinien

Die Ausgangskennlinien sowohl von bipolaren als auch von MOS-Transistoren haben zwar keinen linearen Verlauf (und sie haben verschiedene physikalische Ursachen, obwohl sie sehr ähnlich verlaufen); man erhält jedoch ebenfalls zwei Betriebspunkte für den digitalen Betrieb, in Bild 7.3 bis Bild 7.6 mit (1) und (0) bezeichnet. Stark vereinfacht zeigen im Digitalbetrieb die Transistoren und die mechanistischen Relaiskontakte eine prinzipiell vergleichbare Arbeitsweise:

- Arbeitskontakt \leftrightarrow bipolarer Transistor vom NPN-Typ \leftrightarrow MOS-Transistor mit N-Kanal
- Ruhekontakt \leftrightarrow bipolarer Transistor vom PNP-Typ \leftrightarrow MOS-Transistor mit P-Kanal

7.1.2 Lastwiderstand und Lasttransistor

Neben den aktiven Bauelementen, den Schalttransistoren, benötigt man in Digitalschaltungen noch passive Bauelemente, um gegebenenfalls ein aktives Bauelement zu ersetzen, d.h. um einen Signalpegel zu erzeugen, der zwischen „Lo“ und „Hi“ liegt. In bipolarer Technologie lassen sich ohmsche Widerstände mit vertretbarem Aufwand herstellen (Bild 7.7); in MOS-Technologie dagegen muß als Kompromiß ein Transistor zum passiven Zweipol verdrahtet werden (Bild 7.8).

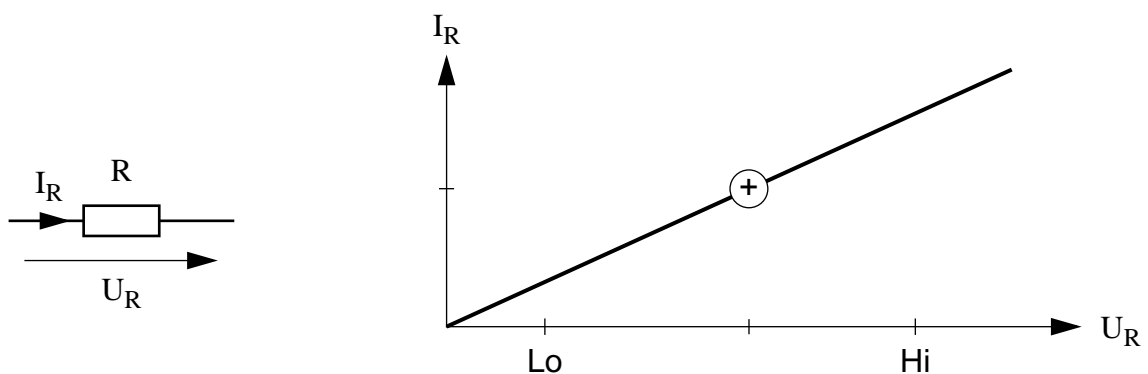


Bild 7.7: Ohmscher Lastwiderstand

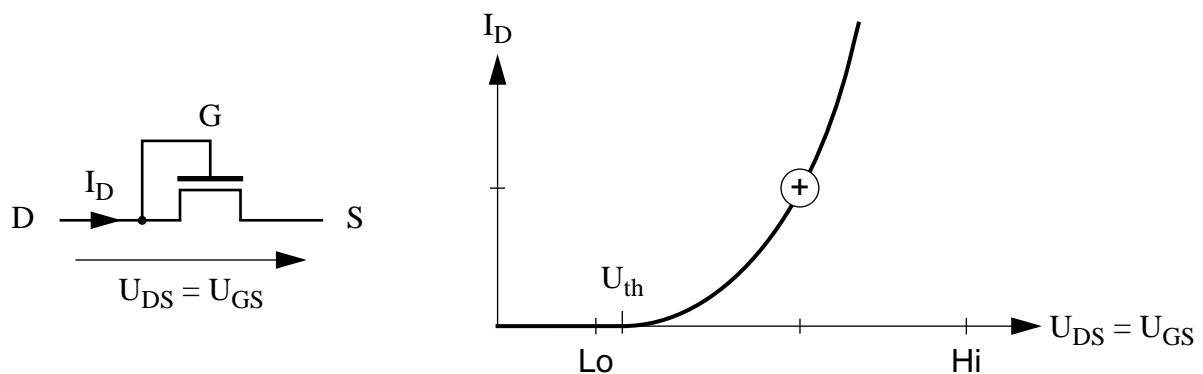


Bild 7.8: MOS-Lasttransistor

7.1.3 Transistoren im Schalterbetrieb

Das nächste Bild zeigt einige einfache Inverterschaltungen in CMOS- bzw. in bipolarer Schaltungstechnik im Vergleich mit Relaisschaltungen mit mechanistischen Kontakten, um ihre Arbeitsweise zu verdeutlichen. Man erkennt, dass die CMOS-Schaltungstechnik im Digitalbetrieb einem Inverter vom *Kontakttyp*, die bipolare Schaltungstechnik einem Inverter vom *Widerstandstyp* entspricht.

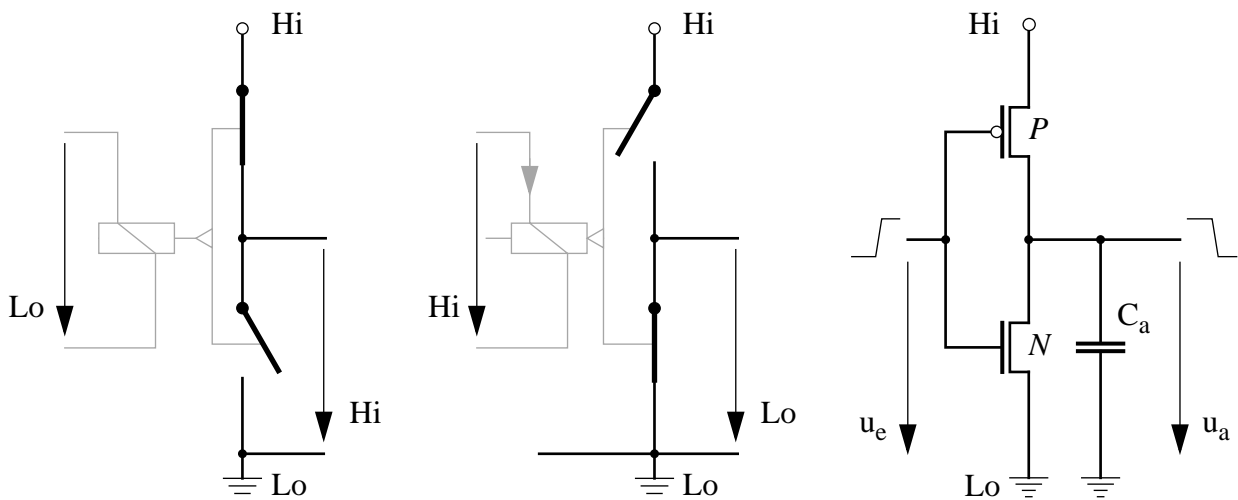


Bild 7.9: Inverterschaltungen vom Kontakttyp. a) in Ruhestellung, b) in Arbeitsstellung, c) Realisierung mit CMOS-Transistoren P und N

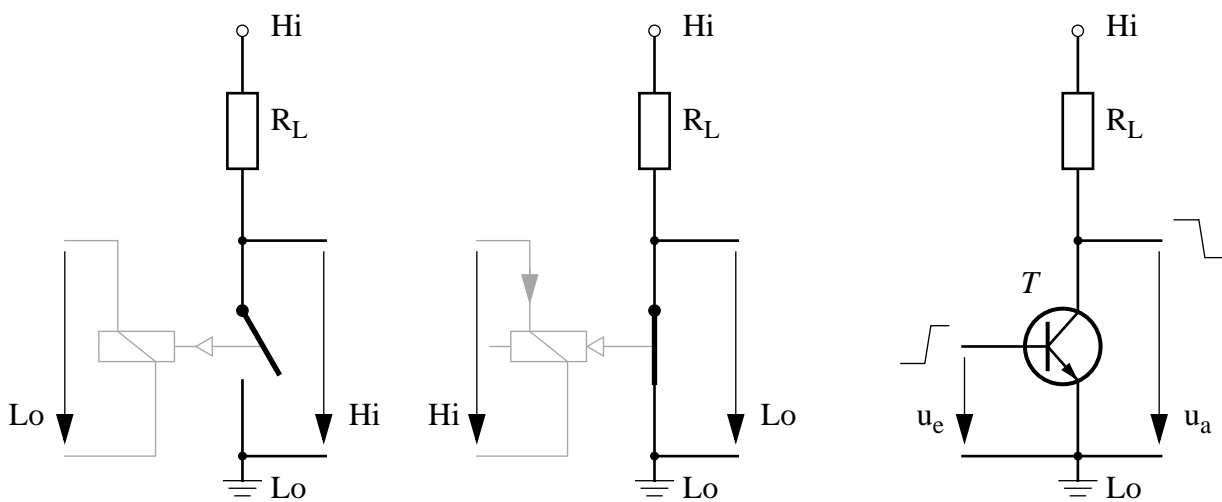


Bild 7.10: Inverterschaltungen vom Widerstandstyp. a) in Ruhestellung, b) in Arbeitsstellung, c) Realisierung mit bipolarem Transistor T

7.2 Elektronische Verstärkung

Da die Ausgangskennlinien sowohl von bipolaren als auch von MOS-Transistoren sehr ähnlich verlaufen, obwohl sie völlig verschiedene physikalische Ursachen haben, werden sie im Folgenden zur Vereinfachung einheitlich und gemeinsam betrachtet.

7.2.1 Verstärkerbetrieb mit kleinen Signalen

Bild 7.11 zeigt zwei einfache Verstärkerschaltungen mit einem bipolarem bzw. einem NMOS-Transistor. Durch den Spannungsteiler R_1 - R_2 wird der „Arbeitspunkt“ der Schaltung eingestellt.

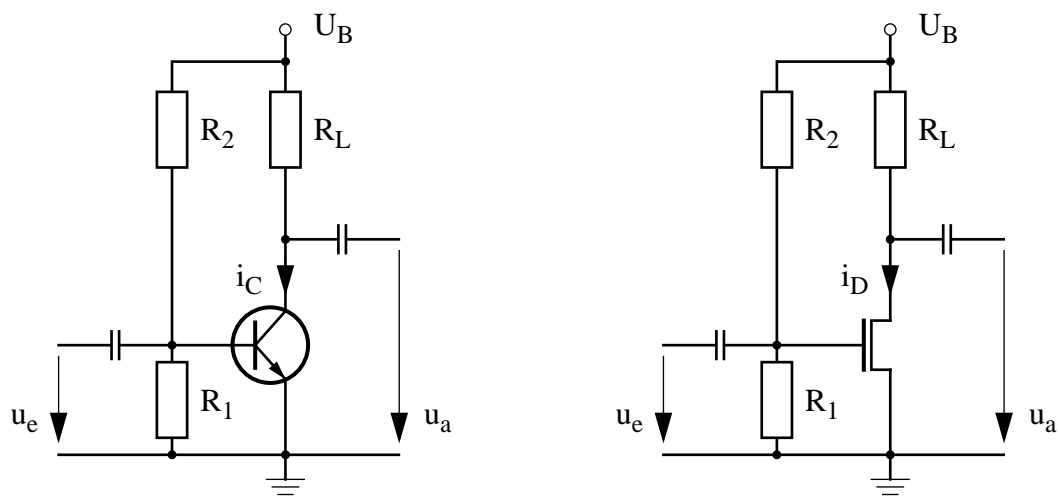


Bild 7.11: Einfache Verstärkerschaltungen; a) mit bipolarem Transistor, b) mit MOS-Transistor.

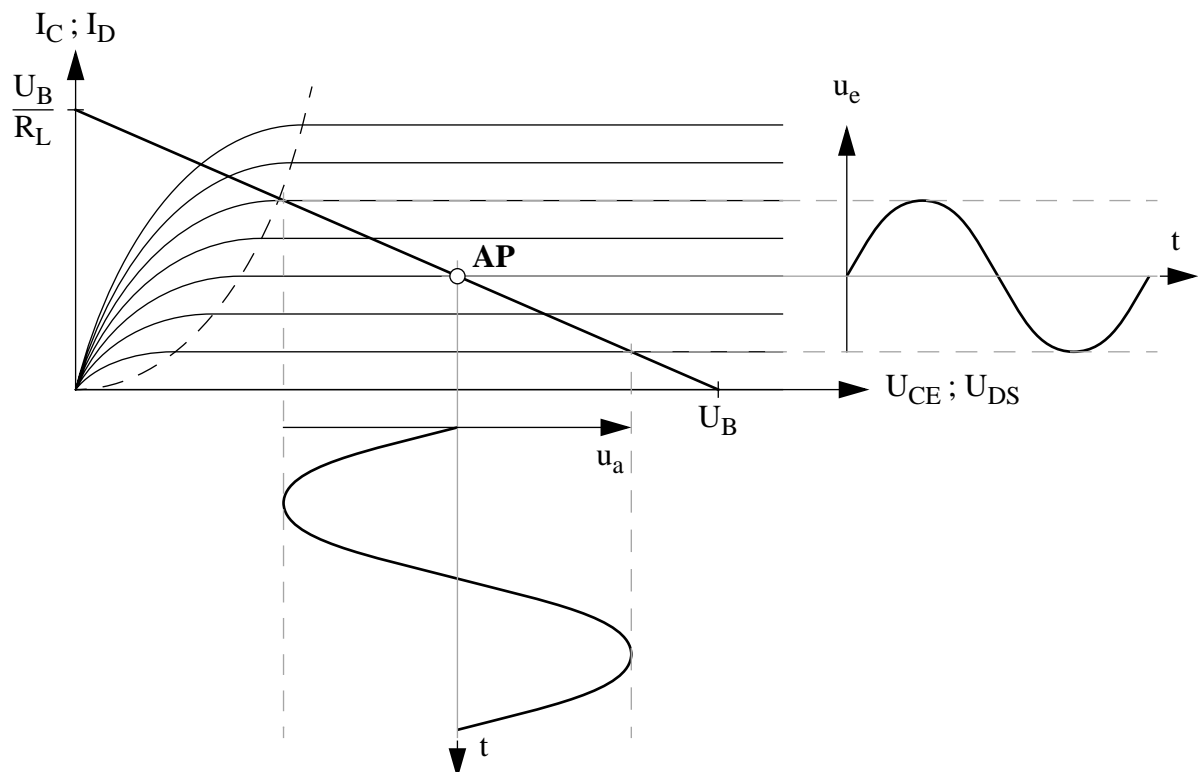


Bild 7.12: Verstärkerbetrieb („Kleinsignalbetrieb“); AP - Arbeitspunkt.

Im Verstärkerbetrieb werden *sinusförmige* Signale verwendet (Bild 7.12). Die Ansteuerung des verstärkenden Transistors erfolgt symmetrisch zu einem mittleren, stabilen Arbeitspunkt (AP), und zwar mit relativ *kleiner* Amplitude, damit die aktuellen Betriebspunkte im linearen Verlauf der Transistorkennlinie verbleiben, um das Signal möglichst unverzerrt zu verstärken („Kleinsignalbetrieb“). Eingangs- und Ausgangssignal schwingen in Gegenphase.

7.2.2 Digitalbetrieb mit großen Signalen

Bild 7.13 zeigt zwei einfache Inverterschaltungen vom Widerstandstyp mit einem bipolarem bzw. einem NMOS-Transistor.

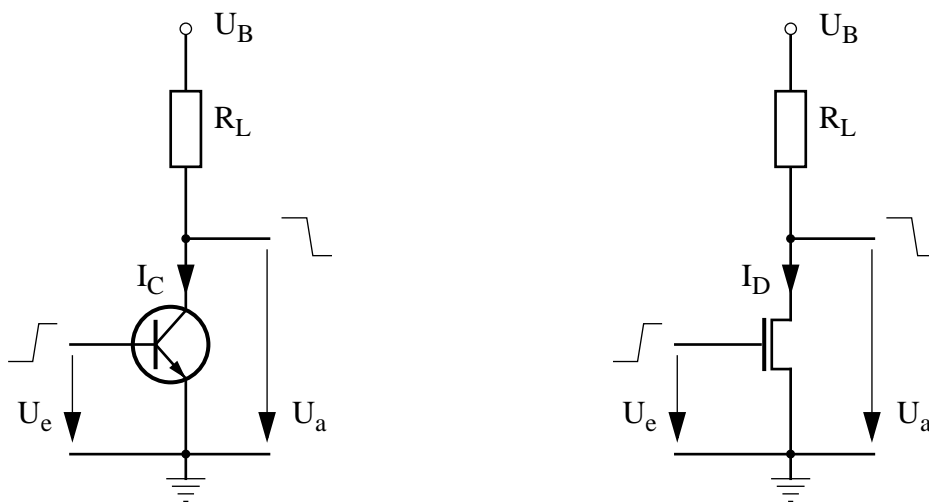


Bild 7.13: Einfache Inverterschaltungen; a) mit bipolarem Transistor, b) mit MOS-Transistor.

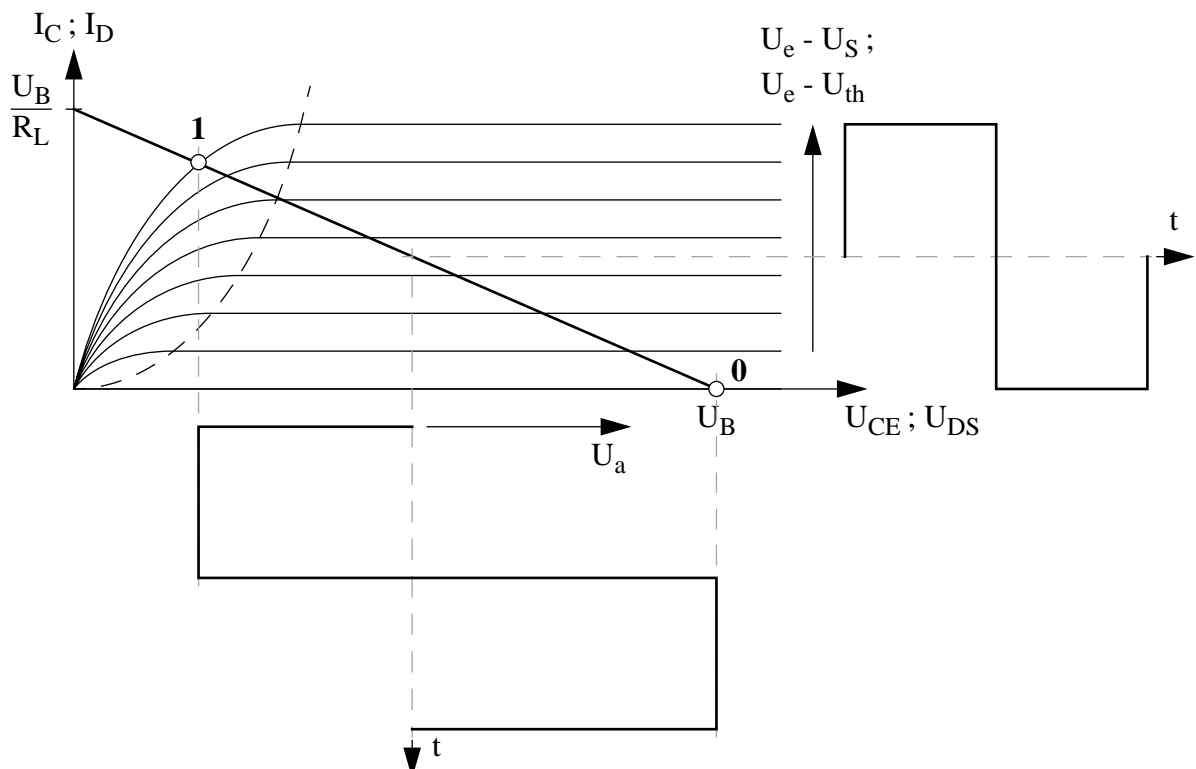


Bild 7.14: Digitaler Impulsbetrieb („Großsignalbetrieb“); (1) (0) - Arbeitspunkte.

Im digitalen Impulsbetrieb werden *rechteckförmige* Signale verwendet (Bild 7.14). Die Ansteuerung des verstärkenden Transistors erfolgt mit relativ *großer* Amplitude, da es im Digitalbetrieb nur auf das Erreichen zweier diskreter Arbeitspunkte (0) bzw. (1) ankommt („Großsignalbetrieb“). Die Schaltung verstärkt das Eingangssignal und invertiert dabei dessen logischen Signalpegel. Man erkennt, dass im Großsignalbetrieb eine Verzerrung des Signals auftritt, die jedoch den Digitalbetrieb nicht stört. Eine hohe Leistungsverstärkung ist in umfangreichen Schaltungen der Informationstechnik unverzichtbar, um Energieverluste bei der Signalverarbeitung und -übertragung auszugleichen.

7.3 Transistoren im Schalterbetrieb

7.3.1 Bipolare Transistoren als Schalter

Der Schalter- oder Großsignalbetrieb eines bipolaren Transistors läßt sich anhand seines Ausgangskennlinienfeldes veranschaulichen, das im folgenden Bild dargestellt ist.

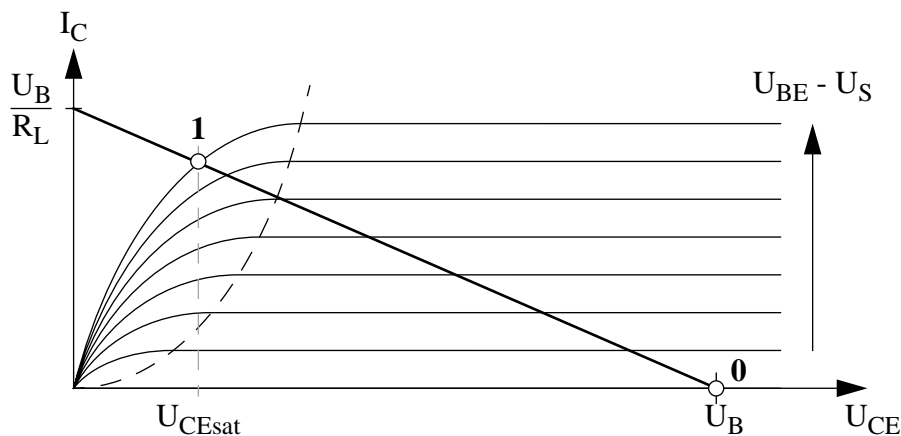


Bild 7.15: Ausgangskennlinienfeld eines bipolaren Transistors vom NPN-Typ

- Liegt an den Eingangsklemmen der Schaltung ein ausreichend niedriger Spannungspegel („Lo“)

Lo: $U_{BE} < [U_S \approx 0,7 \text{ V}]$

Hi: $U_{CE} \approx [U_B > U_S]$

so *sperrt* der Transistor; es fließt nur noch ein vernachlässigbar kleiner Collectorreststrom I_{Ck} . Der Lastwiderstand R_L ist praktisch stromlos, d.h. an ihm fällt keine nennenswerte Spannung ab, so daß sich an den Ausgangsklemmen der hohe Spannungspegel („Hi“) einstellt. Der Arbeitspunkt (0) des Transistors befindet sich im Sperrbereich des Kennlinienfeldes.

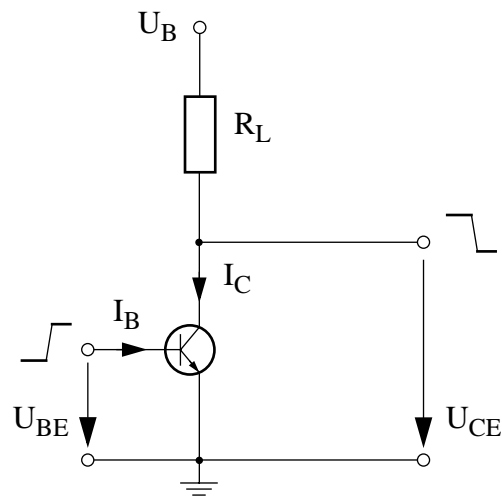


Bild 7.16: Inverterschaltung mit bipolarem Transistor

- Wird die Eingangsspannung auf einen ausreichend hohen Pegel („Hi“) angehoben

Hi: $U_{BE} > [U_s \approx 0,7 \text{ V}]$

Lo: $U_{CEsat} \approx [0,2 \text{ V} < U_s]$

so steigt wegen der Stromverstärkung des Transistors auch sein Collectorstrom I_C an: der Transistor *leitet*. Wäre er ein idealer Kurzschluß, würde am Lastwiderstand R_L die gesamte Betriebsspannung U_B abfallen, d.h. der Maximalwert des Collectorstroms ist $I_C \leq U_B / R_L$. Wegen der Sättigung des Transistors bleibt an den Ausgangsklemmen jedoch die Sättigungsspannung U_{CEsat} bestehen, die niedriger ist als die Schleusenspannung U_s . Der Arbeitspunkt (1) des Transistors befindet sich im Sättigungsbereich des Kennlinienfeldes.

Während des Umschaltens wandert der Arbeitspunkt des Transistors auf der „Arbeitsgeraden“ durch das Kennlinienfeld; ihre Steigung ist durch den Wert des Lastwiderstandes R_L gegeben. Wesentlich für den Betrieb sind die Umschaltzeiten zwischen den beiden Arbeitspunkten. Wärmeverluste im Transistor ergeben sich hauptsächlich während des Umschaltens, da dann sowohl der Collectorstrom I_C als auch die Ausgangsspannung U_{CE} merkbare Werte annehmen. Man ist daher auch aus diesem Grund bestrebt, möglichst kurze Umschaltzeiten zu erreichen. Die Geschwindigkeit, mit der das Umschalten zwischen Sperr- und Sättigungsbereich erfolgt, läßt sich aus dem zeitlichen Verlauf des Collectorstroms I_C erkennen, der im folgenden Bild dargestellt ist.

- Die *Einschaltzeit* setzt sich aus Verzögerungs- und Anstiegszeit zusammen.

Während der Verzögerungszeit t_d bleibt der Transistor zunächst noch im Sperrbereich (0), da die Sperrschicht- sowie die Diffusionskapazität der Basis-Emitterdiode umgeladen werden müssen. Nach Ablauf der Verzögerungszeit t_d sei der Collectorstrom auf 10% seines Maximalwerts angestiegen. Während der anschließenden Anstiegszeit t_r steigt der Collectorstrom annähernd nach einer Exponentialfunktion an, deren Zeitkonstante von der Bauweise des Transistors sowie vom Lastwiderstand R_L abhängt. Man rechnet die Anstiegszeit t_r zwischen 10 . . 90% des Maximalwerts des Collectorstroms.

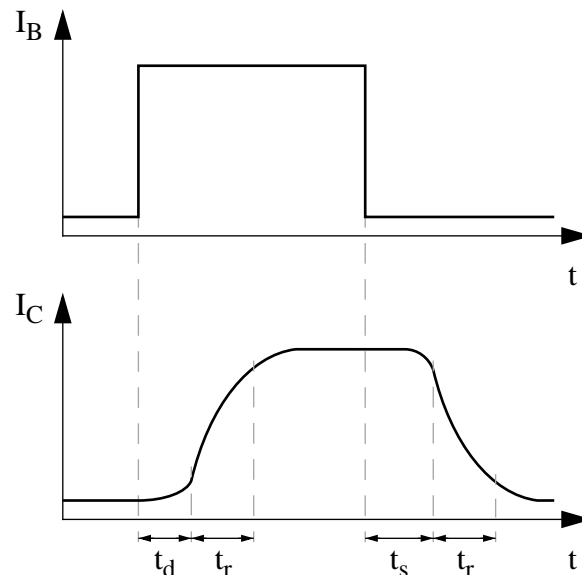


Bild 7.17: Zeitlicher Verlauf von Basis- und Collectorstrom

- Die *Ausschaltzeit* besteht aus Speicher- und Abfallzeit. Während der Speicherzeit t_s bleibt der Transistor zunächst noch im Sättigungsbereich (1), während der Collectorstrom auf 90% seines Maximalwerts abfällt, bis ihre Diffusionskapazität entladen ist. Die Abfallzeit t_f ist durch dieselbe Zeitkonstante wie die Anstiegszeit t_r bestimmt. Man rechnet die Abfallzeit t_f zwischen 90 . . 10% des Maximalwerts des Collectorstroms.

7.3.2 MOS-Transistoren als Schalter

MOS-Transistoren vom Anreicherungstyp können als elektronische Schalter in Digitalschaltungen vorteilhaft verwendet werden, da sie selbstsperrend sind. Durch Ansteuern mit Gate-Source-Spannungen, die abwechselnd unter- bzw. oberhalb der Schwellspannung U_{th} liegen, können sie zwischen dem Betriebszustand ohne und demjenigen mit Kanalbildung umgeschaltet werden. Im Sperrzustand führt der MOS-Transistor praktisch keinen Drainstrom, während eine hohe Drain-Source-Spannung abfallen kann. Im Durchlaßzustand fließt ein hoher Drainstrom bei relativ geringer Drain-Source-Spannung. Da MOS-Transistoren im wesentlichen aus einem MOS-Kondensator bestehen, auf dessen Prinzip ihre Wirkungsweise beruht, sind sie stets mit einer MOS-Kapazität behaftet, die die Schaltgeschwindigkeit entscheidend beeinflusst. Für die anschließenden Betrachtungen werden N-Kanal MOS-Transistoren angenommen.

Der Schalter- oder Großsignalbetrieb eines MOS-Transistors läßt sich ebenfalls anhand seines Ausgangskennlinienfeldes veranschaulichen, das im folgenden Bild dargestellt ist.

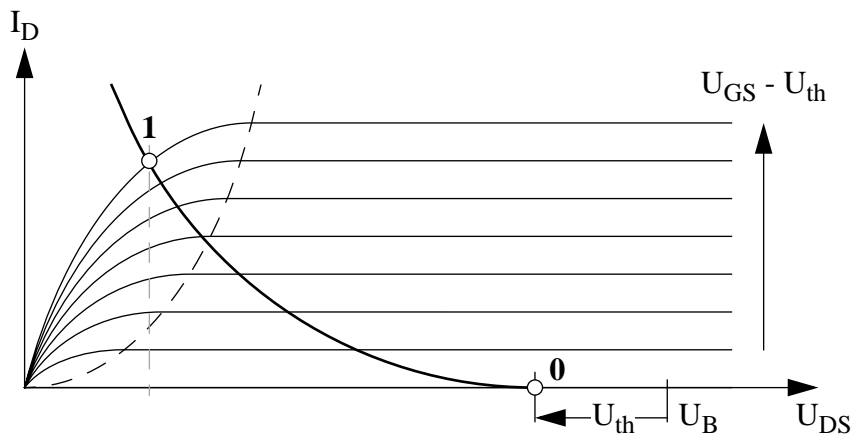


Bild 7.18: Ausgangskennlinienfeld eines MOS-Transistors mit N-Kanal

Der meistverwendete Grundbaustein digitaler MOS-Schaltkreise ist die im Bild rechts dargestellte Inverterschaltung. Der sonst gebräuchliche ohmsche Lastwiderstand muß hier durch einen zum Zweipol verdrahteten MOS-„Lasttransistor“ ersetzt werden, da in hochintegrierter MOS-Technologie keine hochohmigen, d.h. schwach dotierten Gebiete zur Verfügung stehen. Ferner wird der Inverter durch die Eingangskapazität der nachfolgenden Stufe belastet, die bei MOS-Schaltungen prinzipiell nicht vernachlässigbar ist; sie wird hier durch eine Lastkapazität C_A berücksichtigt.

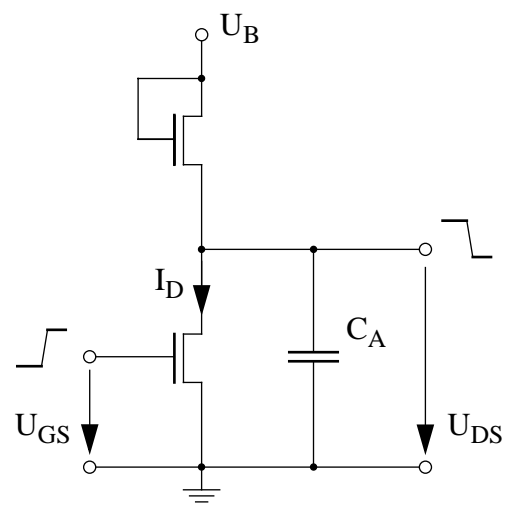


Bild 7.19: Inverterschaltung mit NMOS-Transistoren

- Liegt an den Eingangsklemmen der Schaltung ein ausreichend niedriger Spannungspegel („Lo“)

Lo: $U_{GS} < U_{th} \approx 0,3 \dots 1,0 \text{ V}$

Hi: $U_{DS} \approx U_B - U_{th}$

so *sperrt* der MOS-Transistor. Der Drainstrom ist $I_D = 0$; auch der Lasttransistor ist dann stromlos. Er hat zuvor die Lastkapazität C_A aufgeladen, aber nicht auf die volle Betriebsspannung U_B , da er bereits beim Erreichen der Bedingung $U_{GS} = U_{th}$ sperrt; an den Ausgangsklemmen stellt sich der oben angegebene Spannungspegel „Hi“ ein. Der Arbeitspunkt (0) des MOS-Transistors befindet sich im Sperrbereich des Kennlinienfeldes.

- Wird die Eingangsspannung auf einen ausreichend hohen Pegel angehoben („Hi“)

Hi: $U_{GS} > U_{th}$

Lo: $U_{DS} < U_{th}$

so *leitet* der MOS-Transistor. Im leitenden Zustand wirkt die Inverterschaltung als Spannungsteiler. Will man dabei eine hinreichend niedrige Ausgangsspannung U_{DS} erreichen, so muß der Lasttransistor hochohmig, d.h. mit langem, schmalen Kanal, der Schalttransistor dagegen niederohmig, d.h. mit weitem, kurzem Kanal ausgelegt werden. Der Arbeitspunkt (1) des MOS-Transistors befindet sich dann im linearen Bereich des Kennlinienfeldes.

Die Lage des Arbeitspunkts (1), d.h. die minimale Ausgangsspannung U_{DS} , hängt von der maximalen Eingangsspannung U_{GS} ab.

Sie hängt ferner von der Auslegung der Kanalweiten W und Kanallängen L von Schalt- und Lasttransistor ab. Das Verhältnis $W:L$ des Schalttransistors bestimmt direkt sein Ausgangskennlinienfeld, dasjenige des Lasttransistors die im obigen Ausgangskennlinienfeld eingezeichnete „Arbeitsparabel“. Während des Umschaltens wandert der Arbeitspunkt des MOS-Transistors auf dieser Arbeitsparabel durch das Kennlinienfeld. Die Geschwindigkeit, mit der das Umschalten erfolgt, läßt sich aus dem im Bild rechts dargestellten zeitlichen Verlauf der Ausgangsspannung U_{DS} erkennen:

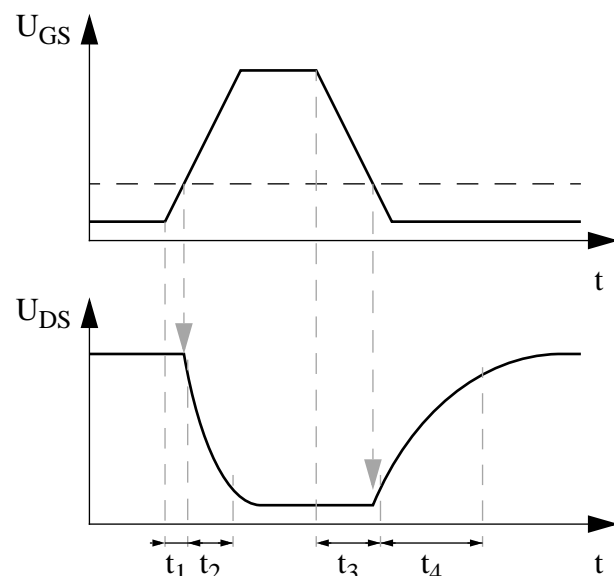


Bild 7.20: Zeitlicher Verlauf der Ein- und Ausgangsspannung

- Die *Einschaltzeit* setzt sich aus zwei Zeitabschnitten zusammen. Liegt die Eingangsspannung U_{GS} zunächst unterhalb der Schwellspannung U_{th} , so ergibt sich die Ausgangsspannung zu $U_{DS} = U_B - U_{th}$. Steigt die Eingangsspannung U_{GS} nicht sprunghaft an, so verstreicht eine gewisse Zeit, bis sie die Schwellspannung U_{th} überschreitet. Dann beginnt der Schalttransistor zu leiten, die Lastkapazität C_A wird entladen und die Ausgangsspannung U_{DS} sinkt ab, nach der Zeitspanne t_1 auf 90% ihres Maximalwerts. Die Entladezeitkonstante hängt vom Kapazitätswert C_A und vom Schalttransistor ab. Da letzterer, wie oben erwähnt, niederohmig ausgelegt wird, geht diese Entladung relativ schnell vonstatten. Man setzt dafür eine Zeit t_2 zwischen 90 . . 10% der maximalen Ausgangsspannung an.

- Auch die *Ausschaltzeit* setzt sich aus zwei Zeitabschnitten zusammen. Sinkt die Eingangsspannung U_{GS} wieder unter den Wert der Schwellspannung U_{th} , so sperrt der MOS-Transistor und die Lastkapazität C_A wird über den Lasttransistor aufgeladen. Fällt die Eingangsspannung U_{GS} nicht sprunghaft ab, so ist eine Verzögerungszeit t_3 zu erwarten, bis die Ausgangsspannung wieder 10% ihres Maximalwerts erreicht. Da der Lasttransistor hochohmig ausgelegt wird (zum einen, um den Spannungsteiler zwischen Last- und Schalttransistor richtig zu dimensionieren, zum andern, um die Verlustleistung klein zu halten), erfolgt die Aufladung der Lastkapazität C_A wesentlich langsamer als deren Entladung. Man setzt dafür eine Zeit t_4 vom 0,1-fachen des Endwerts an, der wieder $U_B - U_{thL}$ beträgt.

Es ist zu beachten, daß die Schaltgeschwindigkeit der Inverterschaltung wesentlich von der geometrischen Auslegung der beteiligten MOS-Transistoren abhängt sowie von der Größe der zu treibenden Lastkapazität und dem zeitlichen Verlauf der Eingangsspannung.

8.1 Realisierung von Schaltfunktionen

Umfangreiche Verknüpfungsnetzwerke zur technischen Informationsverarbeitung entstehen durch Reihen- und Parallelschaltung von standardisierten logischen Grundsaltungen. Ihre binären Ausgangspegel müssen deshalb innerhalb enger Toleranzen gleich den binären Eingangspegeln sein, auch unter Belastung durch die angeschlossenen Schaltkreise. Die zwei unterschiedlichen Pegel seien „Hi“ und „Lo“ genannt. Wie die folgenden Tabellen zeigen, ergeben sich je nach Zuordnung der Pegel zu den binären Werten „0“ und „1“ einer Schaltfunktion zu jeder *Pegeltabelle* zwei logische *Wahrheitstabellen*.

U_{e2}	U_{e1}	U_a	NOR:	e_2	e_1	a	NAND:	e_2	e_1	a
Lo	Lo	Hi	$Hi \leftrightarrow 1$	0	0	1	$Hi \leftrightarrow 0$	1	1	0
Lo	Hi	Lo	$Lo \leftrightarrow 0$	0	1	0	$Lo \leftrightarrow 1$	1	0	1
Hi	Lo	Lo		1	0	0		0	1	1
Hi	Hi	Lo		1	1	0		0	0	1

U_{e2}	U_{e1}	U_a	NAND:	e_2	e_1	a	NOR:	e_2	e_1	a
Lo	Lo	Hi	$Hi \leftrightarrow 1$	0	0	1	$Hi \leftrightarrow 0$	1	1	0
Lo	Hi	Hi	$Lo \leftrightarrow 0$	0	1	1	$Lo \leftrightarrow 1$	1	0	0
Hi	Lo	Hi		1	0	1		0	1	0
Hi	Hi	Lo		1	1	0		0	0	1

Bild 8.1: Pegeltabellen logischer Schaltungen (links);
Wahrheitstabellen in positiver Logik (mitte), in negativer Logik (rechts)

Zur Realisierung von Pegeltabellen wurden unterschiedliche Schaltkreisfamilien in bipolarer und MOS-Technologie entwickelt, deren wichtigste Vertreter in den folgenden Abschnitten vorgestellt werden.

8.2 Logik mit bipolaren Transistoren

Wir wollen uns hier auf Transistoren vom NPN-Typ beschränken. Ferner sollen die Betrachtungen grundsätzlich für positive Logik gelten. Beim Schaltungsentwurf mit bipolaren Transistoren sind drei charakteristische Merkmale besonders zu berücksichtigen:

- Der bipolare Transistor besitzt keinen unendlich großen Eingangswiderstand, d.h. er benötigt unter Umständen einen beträchtlichen *Basisstrom* $I_B > 0$, dessen Anlieferung durch geeignete Anordnung und Dimensionierung der Bauelemente gewährleistet werden muß.
- Eine Silizium-Diode (und damit auch der Basis-Emitter-Übergang eines Silizium-Transistors) benötigt, bevor in Durchlaßrichtung ein merklicher Strom fließt, eine Vorspannung, die höher ist als eine *Schleusenspannung* $U_S \approx 0,7 \text{ V}$. Jeder Einbau einer leitenden Diode oder eines leitenden Basis-Emitter-Übergangs ist daher mit einer Pegelverschiebung um den Betrag dieser Schleusenspannung verbunden, die gegebenenfalls kompensiert werden muß.
- Ein leitender Transistor stellt keinen idealen Kurzschluß dar. Selbst wenn er bis in den Sättigungsbereich gesteuert wird, sinkt seine Collector-Emitter-Spannung nur bis auf eine *Sättigungsrestspannung* $U_{CEsat} \approx 0,2 \text{ V}$ ab.

Daraus ergeben sich für den Digitalbetrieb von bipolaren Transistoren zwei wohlunterschiedene Spannungspegel mit der Schleusenspannung U_S als Trennwert sowie der Restspannung U_{CEsat} als unterer und der Betriebsspannung U_B als oberer Grenze:

$$[U_{CEsat} \approx 0,2 \text{ V}] \leq \mathbf{Lo} < [U_S \approx 0,7 \text{ V}] \quad [U_S \approx 0,7 \text{ V}] < \mathbf{Hi} \leq U_B$$

Folgende Schaltkreisfamilien mit bipolaren Transistoren sollen hier vorgestellt werden:

- Direkt gekoppelte Transistor-Logik (DCTL)
- Dioden-Transistor-Logik (DTL)
- Transistor-Transistor-Logik (TTL)
- Emmittergekoppelte Logik (ECL)

8.2.1 Direkt gekoppelte Transistor-Logik (DCTL)

Eine elementare Verknüpfungsschaltung ist die im folgenden Bild (links) gezeigte Inverterschaltung, die aus einer minimalen Anzahl von Bauelementen besteht. Es handelt sich um eine der digitalen Grundsaltungen, deren Arbeitsweise bereits im vorigen Kapitel besprochen wurde. Hier soll nur kurz das wesentliche wiederholt werden, um die Realisierung der logischen NICHT-Funktion mit bipolaren Transistoren zu verdeutlichen. Man erkennt, daß der NPN-Transistor T in Emitterschaltung betrieben wird.

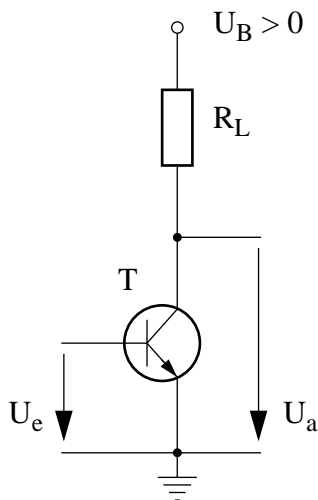


Bild 8.2: Inverterschaltung in DCTL-Technik

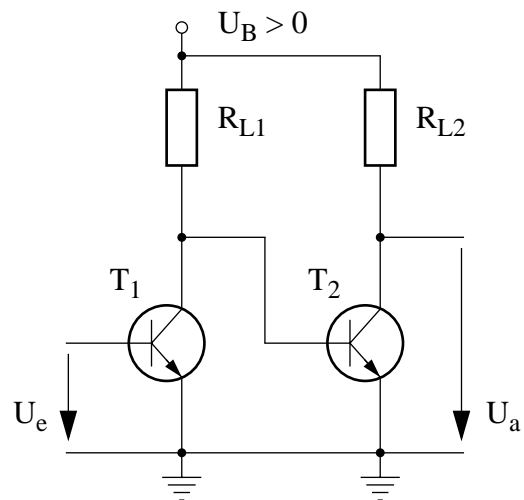


Bild 8.3: Prinzip der direkten Kopplung in DCTL-Technik

Ist die Eingangsspannung der Inverterschaltung niedriger als die Schwellenspannung, d.h. auf dem Wert „Lo“, so sperrt der Schalttransistor T. Die Ausgangsspannung nimmt dann annähernd den Wert der Betriebsspannung $U_B = 5\text{ V}$ an („Hi“), die höher als die Schwellenspannung gewählt werden muß:

$$\text{Lo: } U_e < [U_S \approx 0,7\text{ V}]$$

$$\text{Hi: } U_a \approx U_B > [U_S \approx 0,7\text{ V}]$$

Die sich einstellende Ausgangsspannung U_a kann direkt als Eingangsspannung für gleichartige Schaltungsstufen dienen. Nimmt somit die Eingangsspannung den Wert „Hi“ an, ist sie also höher als die Schwellenspannung, so leitet der Schalttransistor T und sein Collectorpotential nimmt den Wert der Sättigungsrestspannung an („Lo“):

$$\text{Hi: } U_e \approx U_B > [U_S \approx 0,7\text{ V}]$$

$$\text{Lo: } U_a = [U_{CSat} \approx 0,2\text{ V}] < [U_S \approx 0,7\text{ V}]$$

Das Prinzip der direkten Kopplung verdeutlicht das obige Bild (rechts) am einfachen Beispiel zweier hintereinander geschalteter Inverterstufen: Das Ausgangspotential der ersten Stufe ist identisch mit dem Eingangspotential der zweiten. Für die beiden Pegel, die dieses Potential im Digitalbetrieb annehmen kann, ergeben sich etwa folgende Zahlenwerte:

- Leitet Transistor T_1 , so sinkt sein Collectorpotential auf $U_{CEsat} \approx 0,2\text{ V}$ ab, liegt also unterhalb der Schwellenspannung $U_S \approx 0,7\text{ V}$ des Transistors T_2 , so daß dieser gesperrt wird. Der Störabstand

beträgt $\Delta U_{Lo} = U_S - U_{CEsat} \approx 0,5 \text{ V}$. Er kann aufgrund von Fertigungstoleranzen jedoch beträchtlich kleiner ausfallen; dann kann T_2 unter Umständen nicht mehr zuverlässig gesperrt werden.

- Sperrt Transistor T_1 , so muß sein Collectorpotential über $U_S \approx 0,7 \text{ V}$ ansteigen, d.h. es muß $U_B > U_S$ gewählt werden. Um Transistor T_2 bis in den Sättigungsbereich zu steuern, genügt schon ein Wert von $U_B \geq 0,9 \text{ V}$, was einem Störabstand von nur $\Delta U_{Hi} = U_B - U_S \approx 0,2 \text{ V}$ entspricht.

Wählt man anstelle von Transistor T_1 eine Parallelschaltung aus mehreren Transistoren, um die Inverterstufe zu einer Verknüpfungstufe zu erweitern, so ändert sich an der Betriebssicherheit der direkten Kopplung nichts. Ersetzt man dagegen den Transistor T_1 durch eine Reihenschaltung von z.B. drei Transistoren, kann die Ausgangsspannung dieser Schaltung nur auf $3 \cdot U_{CEsat} \approx 3 \cdot 0,2 \text{ V}$ absinken, d.h. der Störabstand beträgt nur noch $\Delta U_{Lo} = U_S - 3 \cdot U_{CEsat} \approx 0,7 \text{ V} - 0,6 \text{ V} \approx 0,1 \text{ V}$, so daß die nachfolgende Stufe, hier Transistor T_2 , nicht mehr sicher gesperrt wird. Deshalb ist eine Parallel- der Reihenschaltung vorzuziehen, wie nachfolgend näher erläutert wird. Ersetzt man also den Schalttransistor T in einer Inverterschaltung (Bild 8.2) durch eine Parallel- oder eine Reihenschaltung mehrerer Transistoren, so erhält man die beiden grundlegenden Verknüpfungsschaltungen in DCTL-Technik ("Direct Coupled Transistor Logic"), wie sie im folgenden Bild dargestellt sind.

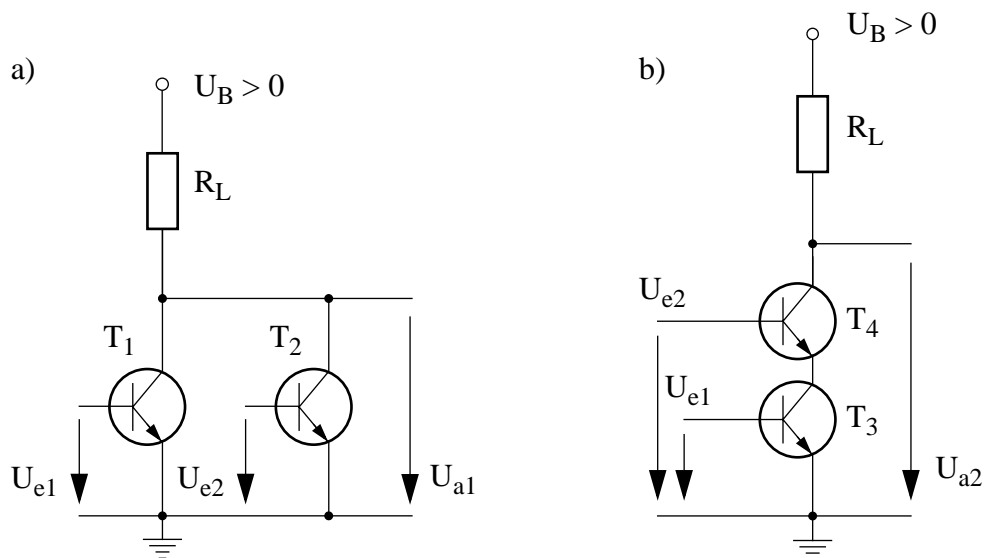


Bild 8.4: Grundsaltungen der DCTL-Technik

U_{e2}	U_{e1}	U_{a1}
Lo	Lo	Hi
Lo	Hi	Lo
Hi	Lo	Lo
Hi	Hi	Lo

U_{e2}	U_{e1}	U_{a2}
Lo	Lo	Hi
Lo	Hi	Hi
Hi	Lo	Hi
Hi	Hi	Lo

Bild 8.5: Realisierung von Schaltfunktionen in DCTL-Technik

a) NOR-Funktion (in positiver Logik)

b) NAND-Funktion (in positiver Logik)

Wird bei einer Parallelschaltung nach Bild 8.4a („NOR-Schaltung“) das Potential *mindestens einer* der Eingangsklemmen über den Pegel der Schleusenspannung $U_S \approx 0,7 \text{ V}$ angehoben, so wird mindestens einer der Transistoren leitend. Es genügt bereits ein geringer Überschuß, um sie in die Sättigung zu treiben; dabei sinkt das Ausgangspotential U_{a1} auf den Wert der Sättigungsrestspannung $U_{CEsat} \approx 0,2 \text{ V}$ ab:

$$\textbf{Hi: } [U_{e1} > U_S] \text{ oder } [U_{e2} > U_S]$$

$$\textbf{Lo: } U_{a1} = U_{CSsat} < U_S \quad (8.1)$$

Wird bei der Parallelschaltung das Potential *aller* Eingangsklemmen auf den Wert der Restspannung $U_{CEsat} \approx 0,2 \text{ V}$ abgesenkt, d.h. unter die Schleusenspannung $U_S \approx 0,7 \text{ V}$, so fließen über den Lastwiderstand R_L nur noch geringe Transistor-Sperrströme sowie die niedrigen Basisströme von an den Ausgang angeschlossenen, gleichartigen Schaltungen, so daß die Ausgangsspannung ansteigt, und zwar praktisch bis auf den Wert der Betriebsspannung U_B :

$$\textbf{Lo: } U_{e1} \text{ und } U_{e2} = U_{CSsat} < U_S ,$$

$$\textbf{Hi: } U_{a1} \approx U_B > U_S \quad (8.2)$$

In positiver Logik realisiert die Parallelschaltung eine NOR-, in negativer eine NAND-Funktion.

Bei einer Reihenschaltung nach Bild 8.4b („NAND-Schaltung“) kann das Ausgangspotential U_{a2} nur dann absinken, wenn alle Transistoren gleichzeitig leiten. Dazu muß das Potential *aller* Eingangsklemmen über den Pegel der Schleusenspannung $U_S \approx 0,7 \text{ V}$ angehoben werden, wobei zu beachten ist, daß die Emitterpotentiale der Transistoren um die Sättigungsspannungen der jeweils darunterliegenden, leitenden Transistoren ansteigen:

$$\textbf{Hi: } [U_{e1} > U_S] \text{ und } [U_{e2} > U_S + U_{CEsat}]$$

$$\textbf{Lo: } U_{a2} = 2 \cdot U_{CEsat} \quad (8.3)$$

Wird bei der Reihenschaltung das Potential *mindestens einer* der Eingangsklemmen auf den Wert der Restspannung $U_{CEsat} \approx 0,2 \text{ V}$ abgesenkt, d.h. unter die Schleusenspannung $U_S \approx 0,7 \text{ V}$, so fließen über den Lastwiderstand R_L nur noch ein geringer Transistor-Sperrstrom sowie die niedrigen Basisströme von an den Ausgang angeschlossenen, gleichartigen Schaltungen, so daß die Ausgangsspannung ansteigt, und zwar praktisch bis auf den Wert der Betriebsspannung U_B :

$$\textbf{Lo: } U_{e1} \text{ oder } U_{e2}$$

$$\textbf{Hi: } U_{a2} \approx U_B > U_S \quad (8.4)$$

In positiver Logik realisiert die Reihenschaltung eine NAND-, in negativer eine NOR-Funktion.

Man erkennt aus den obigen Gleichungen, daß der Potentialpegel, auf den die Ausgangsspannung eines DCTL-Gatters absinken kann, im Fall der Reihenschaltung höher ist als im Fall der Parallelschaltung, so daß letztere einen höheren Störabstand aufweist. Außerdem nimmt bei der Reihenschaltung der minimal erforderliche Eingangsspannungspegel mit der Anzahl der Transistoren zu, was bei der Parallelschaltung nicht der Fall ist. Deshalb wird in der Praxis die Parallelschaltung meist vorgezogen.

8.2.2 Dioden-Transistor-Logik (DTL)

Durch Aneinanderfügen eines passiven Diodengatters und einer verstärkenden Inverterstufe entsteht der im nächsten Bild dargestellte DTL-Schaltkreis („Dioden-Transistor-Logik“). Die Dioden D_1 und D_2 übernehmen, zusammen mit dem Widerstand R_1 , die logische Verknüpfung, in diesem Beispiel eine UND-Funktion. Der Transistor T_A sorgt für die Spannungsverstärkung, was mit einer logischen Negation verbunden ist, so daß insgesamt eine NAND-Verknüpfung entsteht. Die Diode D_B dient in Verbindung mit dem Widerstand R_2 zur Pegelanpassung, wie unten erläutert werden wird.

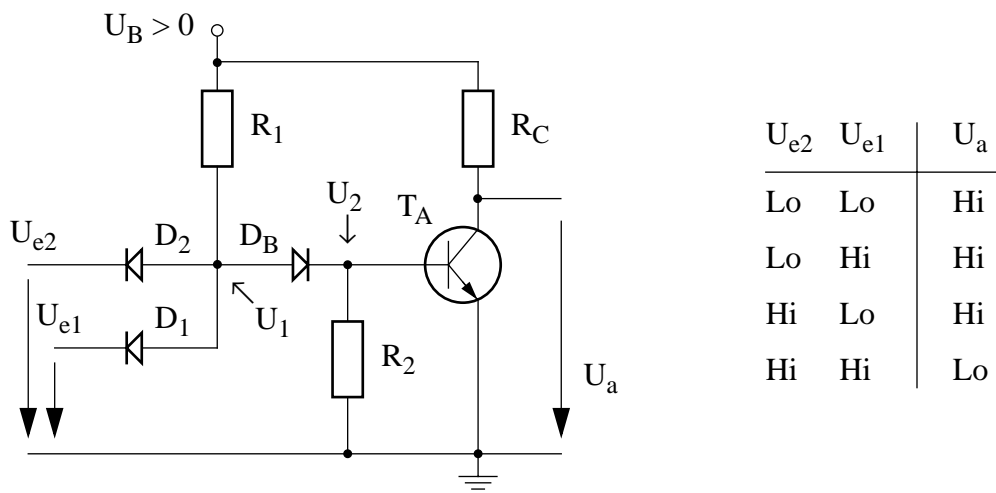


Bild 8.6: DTL-Technik zur Realisierung der NAND-Funktion in positiver Logik

Nur wenn an *alle* Eingangsklemmen gleichzeitig ein hohes Potential angelegt wird („Hi“), das im Digitalbetrieb hier fast den Wert der Betriebsspannung $U_B = 5\text{ V}$ erreicht, so sind alle Eingangsdiode gesperrt. Dann erhält der Transistor T_A über den Widerstand R_1 und die Diode D_B Basisstrom, so daß er leitend wird, wobei er sättigt. Dabei sinkt sein Collectorpotential bis auf den Wert der Sättigungsrestspannung („Lo“) ab:

$$\text{Hi: } U_{e1} \text{ und } U_{e2} \approx U_B > U_S \qquad \text{Lo: } U_a = U_{CEsat} < U_S \qquad (8.5)$$

Im Fall des leitenden Transistors T_A ergibt sich ein relativ kleiner Störabstand ΔU , da bereits ein Überschuß von etwa $0,2\text{ V}$ über dem Wert der Schleienspannung U_S als Basis-Emitter-Spannung genügt, um einen Transistor kräftig zu sättigen:

$$\Delta U_{Hi} = U_B - U_S \geq 0,9\text{ V} - 0,7\text{ V} \approx 0,2\text{ V}. \qquad (8.6)$$

Wird *mindestens eine* der Eingangsklemmen auf den Wert der Sättigungsrestspannung („Lo“) abgesenkt, z.B. durch den leitenden Transistor einer vorhergehenden Stufe, so fließt über die Reihenschaltung aus der zugehörigen Eingangsdiode und dem Widerstand R_1 ein Strom. Bekanntlich fällt an einer leitenden Siliziumdiode wenig mehr als ihre Schleienspannung U_S ab, so daß das gemeinsame Anodenpotential der Dioden ungefähr den Wert U_1 annimmt:

$$\text{Lo: } U_{e1} \text{ oder } U_{e2} = U_{CEsat} < U_S \qquad U_1 = U_{CEsat} + U_S > U_S$$

Man sieht sofort, daß U_1 als Basispotential für den Transistor T_A zu hoch wäre, um ihn zu sperren, wie es die NAND-Funktion in diesem Fall verlangt. Durch den (hochohmigen) Widerstand R_2 wird aber in der Diode D_B ein kleiner Durchlaßstrom eingepreßt, so daß an ihr etwa ihre Schleusenspannung U_S abfällt. Damit nimmt das Basispotential U_2 des Transistors T_A den Wert an:

$$U_2 = U_1 - U_S = U_{CEsat} < U_S$$

Da die Restspannung U_{CEsat} eines leitenden Transistors der vorhergehenden Stufe grundsätzlich kleiner ist als die Schleusenspannung U_S eines Basis-Emitter-Übergangs, wird jetzt Transistor T_A zuverlässig gesperrt. Es ergibt sich ein Störabstand von:

$$\Delta U_{Lo} = U_S - U_2 = U_S - U_{CEsat} \approx 0,7 \text{ V} - 0,2 \text{ V} \approx 0,5 \text{ V}. \quad (8.7)$$

Berücksichtigt man, daß die Schleusenspannungen der beteiligten PN-Übergänge durch Fertigungstoleranzen voneinander abweichen können, so muß mit einem etwas kleineren Störabstand gerechnet werden. Das Collectorpotential U_a des gesperrten Transistors T_A steigt praktisch auf den Wert der Betriebsspannung U_B an:

$$\textbf{Hi: } U_a \approx U_B > U_S \quad (8.8)$$

Liegen mehrere oder alle Eingangsklemmen auf dem niedrigeren Potentialpegel U_{CEsat} , so wird der Transistor T_A selbstverständlich ebenfalls gesperrt. Der DTL-Schaltkreis realisiert somit in positiver Logik eine NAND-, in negativer eine NOR-Funktion.

Die Diode D_B hilft jedoch nicht nur *gleichstrommäßig*, d.h. beim sicheren Sperren von Transistor T_A , sondern auch *wechselstrommäßig* beim schnelleren Abschalten des Transistors: Er ist im leitenden Zustand kräftig gesättigt, d.h. in seinem Basisraum ist eine beträchtliche Ladung gespeichert, was ohne besondere Vorkehrungen beim Abschalten zu einer langen Verzögerungszeit aufgrund des Entladevorgangs führt. Die Sperrschichtkapazität der Diode D_B ist jedoch kurzzeitig in der Lage, Ladung aufzunehmen und damit den Basisraum des Transistors teilweise auszuräumen, so daß sich eine kürzere Speicherzeit und damit ein schnelleres Abschalten ergibt.

Die DTL-Technik eignet sich für integrierte Schaltkreise mittleren Integrationsgrades mit einigen hundert Bauelementen pro Chip ("Medium Scale Integration", MSI), da die Dioden und der Transistor in dieser Technologie relativ wenig Platz benötigen. Vor allem können die drei Dioden in einer gemeinsamen „Isolationswanne“ untergebracht werden, da sie eine gemeinsame Anode besitzen.

8.2.3 Transistor-Transistor-Logik (TTL)

Eine Weiterentwicklung der DTL-Technik führt auf die Schaltkreisfamilie der TTL-Technik („Transistor-Transistor-Logik“), deren Grundsaltkreis das nächste Bild zeigt. Man ersetzt im Prinzip die Dioden des DTL-Schaltkreises durch einen Transistor T_E mit entsprechend vielen Emitttern („Multiemitter-Transistor“), um dessen Stromverstärkung vorteilhaft auszunützen.

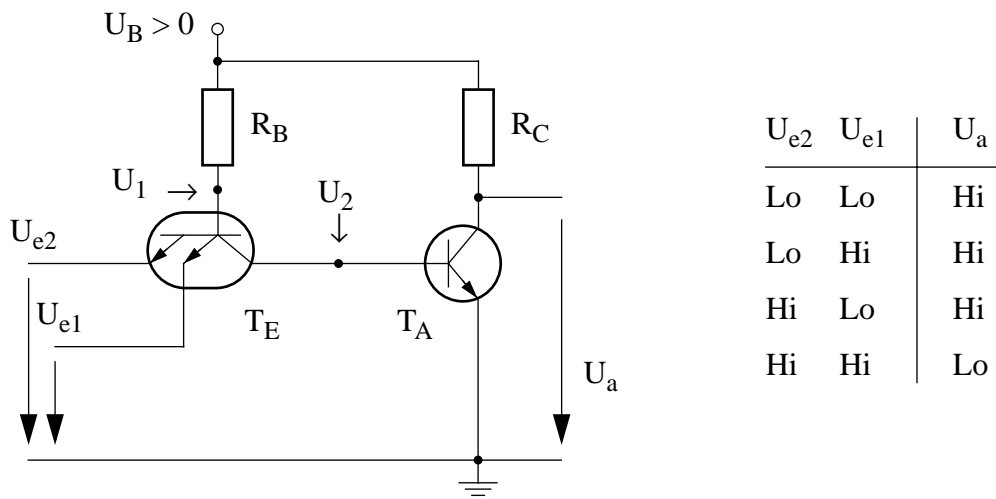


Bild 8.7: TTL-Technik zur Realisierung der NAND-Funktion in positiver Logik

Liegen *alle* Eingangsklemmen gleichzeitig auf hohem Potential („Hi“), das im Digitalbetrieb fast den Wert der Betriebsspannung von $U_B = 5\text{ V}$ erreicht, so sind alle Basis-Emitter-Übergänge des Eingangstransistors T_E gesperrt. Der Ausgangstransistor T_A bezieht dann Basisstrom über den Widerstand R_B und den leitenden Basis-Collector-Übergang von Transistor T_E (der damit invers betrieben wird). Durch den Basisstrom wird Transistor T_A bis in den Sättigungsbereich gesteuert, wobei sein Collectorpotential auf den Wert der Sättigungsrestspannung („Lo“) absinkt:

$$\text{Hi: } U_{e1} \text{ und } U_{e2} \approx U_B > U_S \quad \text{Lo: } U_a = U_{CEsat} < U_S \quad (8.9)$$

Wird *mindestens eine* der Eingangsklemmen auf den Wert der Sättigungsrestspannung („Lo“) abgesenkt, z.B. durch den leitenden Ausgangstransistor einer vorhergehenden Stufe, so wird der zugehörige Basis-Emitter-Übergang von Transistor T_E leitend, wobei sein Basispotential ungefähr auf den Wert U_1 absinkt:

$$\text{Lo: } U_{e1} \text{ oder } U_{e2} = U_{CEsat} < U_S \quad U_1 = U_{CEsat} + U_S > U_S$$

Transistor T_E bekommt über den Widerstand R_B Basisstrom und sättigt. Das bedeutet, daß auch sein Basis-Collector-Übergang in Durchlaßrichtung gepolt wird, so daß das Basispotential von Transistor T_A ebenfalls absinkt, und zwar auf etwa:

$$U_2 = U_1 - U_S = U_{CEsat} < U_S$$

Damit wird der Ausgangstransistor T_A gesperrt und sein Collectorpotential U_a steigt praktisch auf den Wert der Betriebsspannung U_B an:

$$\text{Hi: } U_a \approx U_B > U_S \quad (8.10)$$

Man stellt fest, daß bezüglich der Störabstände für die TTL-Technik dasselbe wie für die DTL-Technik gilt. In beiden Schaltungstechniken kann der Störabstand erhöht werden, indem man der Basis des Ausgangstransistors weitere Dioden in Reihe vorschaltet.

Liegen mehrere oder alle Eingangsklemmen auf dem niedrigeren Potentialpegel U_{CEsat} , so wird der Ausgangstransistor T_A selbstverständlich ebenfalls gesperrt. Der TTL-Schaltkreis realisiert somit in positiver Logik eine NAND-, in negativer eine NOR-Funktion.

Besonders die TTL-Technik ist für integrierte Schaltkreise mittlerer Integrationsdichte (MSI) geeignet: Wie das folgende Bild zeigt, benötigt ein Multiemitter-Transistor nur eine einzige Isolationswanne, da er trotz getrennter Eitterdiffusionen eine gemeinsame Basisdiffusion besitzt, was besonders viel Platz spart. Die Basis-Emitter-Übergänge des Multiemitter-Transistors entsprechen den Eingangsdiolen der DTL-Technik; sein Basis-Collector-Übergang bildet die zur Potentialverschiebung benötigte Diode D_B . Die Wirkungsweise der TTL-Schaltungstechnik ist gleichstommäßig vergleichbar zur DTL-Technik; wechselstrommäßig besitzt sie jedoch entscheidende Vorteile, wie anschließend gezeigt werden wird.

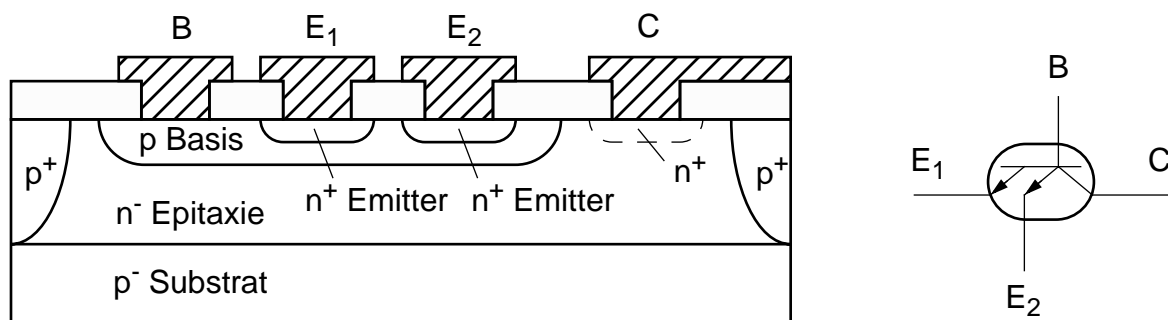


Bild 8.8: Multi-Emitter-Transistor; a) Querschnitt, b) Schaltsymbol

8.2.4 Emittergekoppelte Logik (ECL)

Im Gegensatz zu den bisher besprochenen Schaltkreisfamilien enthält die ECL-Technik („Emitter Coupled Logic“) keine gesättigten Transistoren, so dass die damit verbundenen Speicherzeiten, die sich als Abschaltverzögerungen äußern, von vornherein vermieden werden. Die extrem kurzen Signallaufzeiten durch eine ECL-Stufe sind der Hauptvorteil dieser Schaltungstechnik. Man erkaufte ihn mit einer zusätzlichen Spannungsquelle, die ein Referenzpotential liefern muss.

Den grundlegenden ECL-Schaltkreis zeigt das folgende Bild. Die Schaltung besteht aus mindestens zwei Transistoren T_1 und T_2 , die die logische Verknüpfung von zwei Eingangsgrößen U_{e1} und U_{e2} vornehmen, und einem Transistor T_3 , an dessen Basis das Referenzpotential U_{ref} angelegt wird.

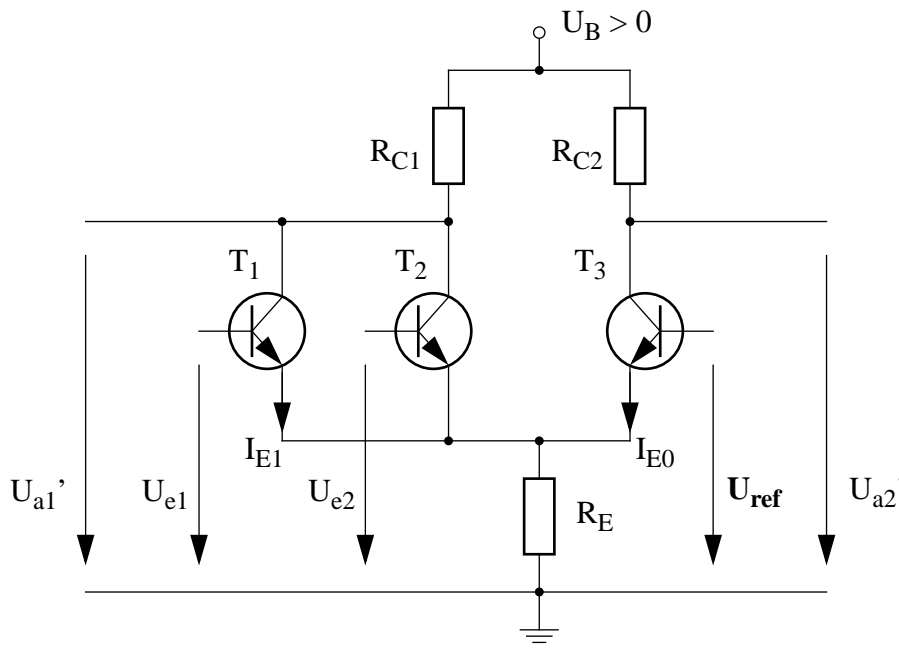


Bild 8.9: Grundsaltkreis in ECL-Technik

Liegen *alle* Eingangsklemmen (hier: e_1 und e_2) auf einem Potential, das niedriger ist als das Referenzpotential:

$$\text{Lo: } U_{e1} \text{ und } U_{e2} < U_{ref}, \quad (8.11)$$

so sind alle Eingangstransistoren (hier: T_1 und T_2) gesperrt, während der Referenztransistor T_3 leitet. Sein Emittterstrom ist gegeben durch den gemeinsamen Emittterwiderstand R_E , wobei zu beachten ist, dass an der Basis-Emittter-Diode eines leitenden Siliziumtransistors praktisch die Schleusenspannung $U_S \approx 0,7 \text{ V}$ abfällt:

$$I_{E0} = (U_{ref} - U_S) / R_E \quad (8.12)$$

Wird *mindestens eine* der Eingangsklemmen (hier: e_1 oder e_2) auf ein Potential angehoben, das aus physikalischen Gründen nur um einen kleinen Betrag ΔU , d.h. wenige Millivolt höher als das Referenzpotential zu sein braucht:

$$\text{Hi: } U_{e1} \text{ oder } U_{e2} = [U_{ref} + \Delta U], \quad (8.13)$$

so übernimmt der zugehörige Eingangstransistor (hier: T_1 oder T_2) innerhalb weniger Picosekunden den durch R_E fließenden Strom und der Referenztransistor T_3 sperrt. Nehmen wir an, dass der Eingangstransistor T_1 leitet, so gilt für seinen Emittterstrom:

$$I_{E1} = (U_{e1} - U_S) / R_E = (U_{ref} + \Delta U - U_S) / R_E = I_{E0} + \Delta U / R_E \approx I_{E0} \quad (8.14)$$

Da wie erwähnt $\Delta U \ll U_{ref}$ ist, bleibt der Strom durch R_E beim Umschalten praktisch konstant: Der gemeinsame Emittterwiderstand R_E wirkt wie eine „Stromquelle“. Deren Strom fließt, abhängig von den Eingangsspannungspegeln, entweder über den Referenztransistor T_3 oder über die Eingangstransistoren (hier: T_1 und/oder T_2). Im letzteren Fall sinkt deren Collectorpotential auf einen Wert ab, der wie folgt gegeben ist:

$$U_{a1}' = U_B - I_{E1} \cdot R_{C1} > [U_{e1} = U_{ref} + \Delta U] \quad (8.15)$$

Diese Gleichung zeigt, dass man durch die Wahl der Betriebsspannung und eine geeignete Dimensionierung der Widerstände erreichen kann, dass das Collectorpotential U_{a1}' eines leitenden Eingangstransistors höher ist als sein Basispotential U_{e1} , was bedeutet, dass seine Basis-Collector-Diode gesperrt bleibt: Der Transistor sättigt nicht. Man könnte zum Beispiel wählen:

$$U_{a1}' = U_{ref} + U_S / 2 > [U_{e1} = U_{ref} + \Delta U] > U_{ref} \quad (8.16)$$

Da der Referenztransistor T_3 im betrachteten Fall sperrt, steigt sein Collectorpotential praktisch bis auf den Wert der Betriebsspannung an, so dass mit Gl. (8.15) gilt:

$$U_{a2}' \approx U_B > [U_{e1} + I_{E1} \cdot R_{C1} = U_{ref} + \Delta U + I_{E1} \cdot R_{C1}] > U_{ref} \quad (8.17)$$

Man erkennt aus diesen beiden Gleichungen, dass *beide* Collectorpotentiale U_{a1}' und U_{a2}' höher als die Referenzspannung sind, also für eine direkte Ansteuerung nachfolgender ECL-Schaltungsstufen ungeeignet sind. Doch schließt man an die Collectoren der Eingangstransistoren (hier: T_1 und T_2) bzw. des Referenztransistors T_3 je einen „Emitterfolger“ wie im folgenden Bild an, so ergibt sich die erforderliche Potentialverschiebung, wie nachfolgend erläutert wird.

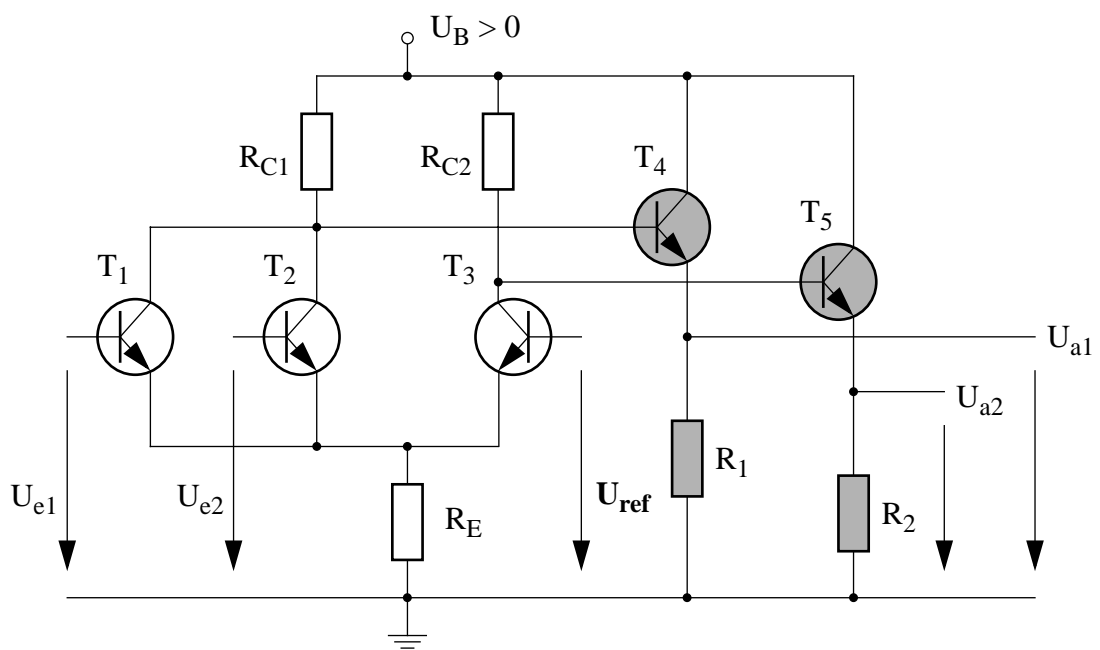


Bild 8.10: Pegelanpassung der NOR- und OR-Ausgänge in ECL-Technik

Die Collectorpotentiale U_{a1}' und U_{a2}' werden je um den Wert der Schleusenspannung $U_S \approx 0,7 \text{ V}$ der Basis-Emitter-Dioden der beiden Ausgangstransistoren T_4 und T_5 abgesenkt:

$$\textbf{Lo: } U_{a1} = U_{a1}' - U_S = (U_{\text{ref}} + U_S / 2) - U_S < U_{\text{ref}} \quad (8.16) \rightarrow \quad (8.18)$$

$$\textbf{Hi: } U_{a2} = U_{a2}' - U_S = U_B - U_S > U_{\text{ref}} \quad (8.17) \rightarrow \quad (8.19)$$

Die beiden Gleichungen zeigen, dass die endgültigen Ausgangspotentiale U_{a1} und U_{a2} beidseits des Referenzpotentials liegen, also logisch komplementär zueinander sind. Diese beiden digitalen Spannungspegel ergeben sich, falls einer, mehrere oder alle Eingangstransistoren leiten. Das bedeutet: Ausgang a_1 liefert in positiver Logik eine NOR-, Ausgang a_2 eine OR-Verknüpfung der Eingangssignale.

Neben den Vorteilen extrem kurzer Umschaltzeiten und der Leistungsverstärkung durch die Ausgangsstufe wird damit ein weiterer Vorteil der ECL-Technik deutlich: Zur Vereinfachung komplexer Verknüpfungsnetzwerke ist es nützlich, gleichzeitig die direkte und die invertierte Schaltfunktion zur Verfügung zu stellen. Das ist vor allem bei der Adressierung von Speicherschaltungen von großer Bedeutung. - Modifikationen der ECL-Grundsaltung beziehen sich insbesondere auf eine Verbesserung der Stromquelle, die im hier gezeigten Ausführungsbeispiel durch einen einfachen Emitterwiderstand R_E realisiert wird.

8.3 Logik mit MOS-Transistoren

MOS steht für “Metal-Oxide-Silicon“, d.h. für Bauelemente mit MOS-Schichtstruktur und ihre Verwendung in mikroelektronischen Schaltungen. Wir wollen uns hier auf den Anreicherungstyp mit isolierter Gate-Elektrode beschränken. Ferner sollen die Betrachtungen grundsätzlich für positive Logik gelten. Beim Schaltungsentwurf mit MOS-Transistoren sind drei charakteristische Eigenschaften besonders zu berücksichtigen:

- Für MOS-Schaltungen in integrierter Technologie ist typisch, daß nicht nur die aktiven sondern auch *passive Bauelemente* durch MOS-Transistoren realisiert werden, weil bei MOS-Fertigungsprozessen die Herstellung hochohmiger Widerstände problematisch ist, da nur niedrige Schichtwiderstände zur Verfügung stehen.
- Ferner ist zu berücksichtigen, daß ein MOS-Transistor vom Anreicherungstyp erst leitet, wenn die Gate-Source-Spannung den Wert einer *Schwellschpannung* U_{th} (“threshold voltage“) überschreitet. Somit ist der Einbau eines MOS-Transistors, ob als aktives oder als passives Bauelement, stets mit einer Pegelverschiebung um den Betrag dieser Schwellspannung verbunden, was bei Bedarf durch geeigneten Einbau weiterer MOS-Transistoren bzw. durch entsprechende Erhöhung der Betriebsspannung ausgeglichen werden muß.
- Schließlich sollte noch beachtet werden, daß jeder MOS-Transistor im Prinzip einen *MOS-Kondensator* von nicht zu vernachlässigender Kapazität darstellt, die bei Schaltvorgängen aufgeladen bzw. entladen werden muß.

Letztere Eigenart ist aber nicht nur störend. Vielmehr ist ein MOS-Transistor in der Lage, Ladung zu speichern und damit seinen augenblicklichen Betriebszustand kurzzeitig beizubehalten, auch wenn die ansteuernden Schaltkreise bereits einen anderen Zustand angenommen haben. Dieser Effekt kann ausgenützt werden, um neben den statischen Verknüpfungsschaltungen, die mit Gleichspannung gespeist werden, auch dynamische zu entwickeln, die von Impulsgeneratoren betrieben werden. Dadurch werden Gleichströme vermieden, was zu extrem niedrigen Verlustleistungen führt. Aus diesen Gründen ermöglicht vor allem die dynamische MOS-Logik den Aufbau höchstintegrierter Schaltungen (“Very Large Scale Integration“, VLSI).

Folgende Schaltkreisfamilien mit MOS-Transistoren sollen hier vorgestellt werden:

- Statische NMOS-Logik
- Statische PMOS-Logik
- Statische CMOS-Logik
- Dynamische Domino-Logik

8.3.1 Statische NMOS-Logik

Für den Digitalbetrieb von MOS-Transistoren vom N-Kanaltyp („NMOS-Transistoren“) gelten zwei wohlunterschiedene Spannungspegel mit der Schwellspannung U_{thN} als Trennwert sowie einer oberen Grenze, die u.a. von der Betriebsspannung von $U_B > 0$, und einer unteren Grenze U_u , deren Wert wie nachfolgend gezeigt von der Dimensionierung der Schaltung abhängt:

$$U_u \leq \mathbf{Lo} < U_{thN}$$

$$U_{thN} < \mathbf{Hi} \leq (U_B - U_{thN})$$

Es gelten folgende Zahlenwerte:

$$U_B \approx +2,5 \text{ V}$$

$$U_{thN} \approx +0,3 \dots +1,0 \text{ V}$$

Das Bild rechts zeigt eine elementare Inverterschaltung mit NMOS-Transistoren; sie realisiert die logische NICHT-Funktion. Der Inverter besteht aus einem (aktiven) Schalttransistor N_1 , der in Source-schaltung betrieben wird, und einem zum Zweipol verdrahteten (passiven) Lasttransistor N_L , da in dieser Technologie bekanntlich keine hochohmigen Widerstände hergestellt werden können. Ferner darf die Eingangskapazität C_a der gleichartigen MOS-Schaltkreise, die von der Inverterschaltung angesteuert werden, hier nicht vernachlässigt werden.

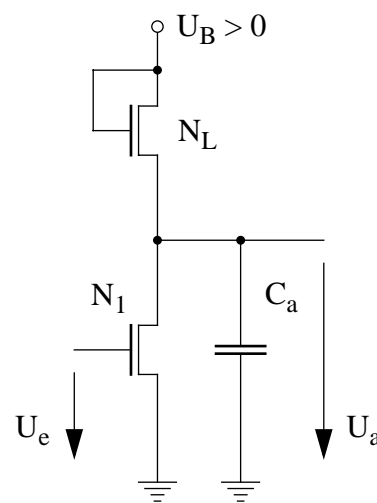


Bild 8.11: Inverterschaltung in statischer NMOS-Logik

Legt man an die Eingangsklemme des NMOS-Inverters einen Potentialpegel, der niedriger liegt als der Wert der Schwellspannung des N-Kanals („Lo“), so sperrt der Schalttransistor N_1 . Die Lastkapazität C_a wird dann über den leitenden Lasttransistor N_L aufgeladen („Hi“):

$$\mathbf{Lo:} \quad U_e < U_{thN}$$

$$\mathbf{Hi:} \quad U_a = (U_B - U_{thN}) > U_{thN}$$

Sobald die Ausgangsspannung auf den Pegel „Hi“ angestiegen ist, erreicht die Gate-Source-Spannung des bisher leitenden Lasttransistors N_L den Wert der Schwellspannung, d.h. er sperrt, so daß sein Sourcepotential U_a nicht weiter ansteigen kann. Der sich einstellende Ausgangspegel U_a kann direkt als Eingangspegel für gleichartige Schaltungsstufen dienen. Nimmt somit der Eingangspegel den Wert „Hi“ an, liegt er also höher als die Schwellspannung, so leitet der Schalttransistor N_1 und entlädt die Lastkapazität C_a bis auf den Pegel „Lo“, der durch das Spannungsteilerverhältnis aus dem Schalttransistor N_1 und dem Lasttransistor N_L gegeben ist. Dabei muß N_1 niederohmig sein, d.h. sein Kanalbereich muß breit und kurz ausgelegt werden, N_L dagegen hochohmig, d.h. lang und schmal, da für den Pegel „Lo“ gelten muß:

$$\mathbf{Hi:} \quad U_e = (U_B - U_{thN}) > U_{thN}$$

$$\mathbf{Lo:} \quad U_a = U_{DS1} < U_{thN}$$

Ersetzt man den Schalttransistor N_1 in der Inverterschaltung (Bild 8.11) durch eine Parallel- oder eine Reihenschaltung mehrerer Schalttransistoren, so erhält man die beiden grundlegenden Verknüpfungsschaltungen in statischer NMOS-Technik, wie sie im folgenden Bild dargestellt sind.

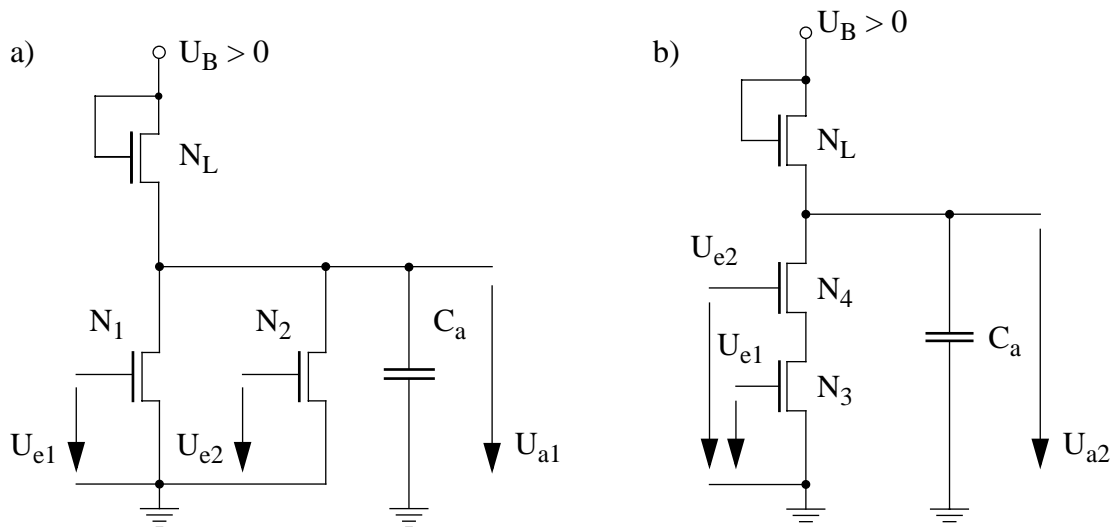


Bild 8.12: Grundsaltungen der statischen NMOS-Logik

U_{e2}	U_{e1}	U_{a1}	U_{e2}	U_{e1}	U_{a2}
Lo	Lo	Hi	Lo	Lo	Hi
Lo	Hi	Lo	Lo	Hi	Hi
Hi	Lo	Lo	Hi	Lo	Hi
Hi	Hi	Lo	Hi	Hi	Lo

Bild 8.13: Realisierung von Schaltfunktionen in statischer NMOS-Logik

a) NOR-Funktion (in positiver Logik)

b) NAND-Funktion (in positiver Logik)

Wird bei einer NMOS-Parallelschaltung nach Bild 8.12a („NOR-Schaltung“) das Potential von *mindestens einer* der Eingangsklemmen über den Pegel der Schwellspannung angehoben („Hi“), so wird mindestens einer der Schalttransistoren leitend. Dadurch entlädt sich die Lastkapazität C_a über mindestens einen der niederohmig ausgelegten Schalttransistoren relativ rasch bis auf den Pegel „Lo“, der durch das Spannungsteilverhältnis aus den jeweils leitenden Schalttransistoren und dem Lasttransistor N_L gegeben ist, das so dimensioniert werden muß, daß das Ausgangspotential U_{a1} niedriger als die Schwellspannung wird:

$$\text{Hi: } [U_{e1} > U_{thN}] \text{ oder } [U_{e2} > U_{thN}]$$

$$\text{Lo: } U_{a1} = U_{DS1..2} < U_{thN} \quad (8.20)$$

Wird bei der Parallelschaltung das Potential *aller* Eingangsklemmen unter den Pegel der Schwellspannung abgesenkt („Lo“), so sperren alle Schalttransistoren und der hochohmig ausgelegte Lasttransistor N_L lädt die Lastkapazität C_a wie bei der Inverterschaltung relativ langsam auf den Pegel „Hi“ auf:

Lo: $U_{e1} \text{ und } U_{e2} < U_{thN}$,

Hi: $U_{a1} = (U_B - U_{thN}) > U_{thN}$ (8.21)

In positiver Logik realisiert die NMOS-Parallelschaltung somit eine NOR-, in negativer eine NAND-Funktion.

Bei einer NMOS-Reihenschaltung nach Bild 8.12b („NAND-Schaltung“) kann die Lastkapazität C_a nur dann entladen werden, wenn alle Schalttransistoren gleichzeitig leiten. Dazu muß das Potential *aller* Eingangsklemmen über den Pegel der Schwellspannung angehoben werden. Dabei ist zweierlei zu beachten: Zum einen verändern sich die Source-Potentiale der Schalttransistoren N_5 und N_6 um die Werte der Drain-Source-Spannungen der jeweils darunterliegenden Schalttransistoren; zum anderen hängt die Schwellspannung von der Source-Body-Vorspannung und damit vom aktuellen Source-Potential ab. Damit gilt für die anzulegenden Eingangsspannungen („Hi“) und die sich einstellende Ausgangsspannung („Lo“):

Hi: $[U_{e1} > U_{thN3}] \text{ und } [U_{e2} > U_{thN4} + U_{DS3}]$

Lo: $U_{a2} = U_{DS4} + U_{DS3} < U_{thN}$ (8.22)

Wird bei der Reihenschaltung das Potential *mindestens einer* Eingangsklemme unter den Pegel der Schwellspannung abgesenkt („Lo“), so sperrt mindestens einer der Schalttransistoren und der hochohmig ausgelegte Lasttransistor N_L lädt die Lastkapazität C_a wie bei der Inverterschaltung relativ langsam auf den Pegel „Hi“:

Lo: $U_{e1} \text{ oder } U_{e2} < U_{thN}$

Hi: $U_{a2} = (U_B - U_{thN}) > U_{thN}$ (8.23)

In positiver Logik realisiert die NMOS-Reihenschaltung somit eine NAND-, in negativer eine NOR-Funktion.

Man erkennt aus einem Vergleich von Gl.(8.22) mit Gl.(8.20), daß der Pegel, auf den die Ausgangsspannung U_a eines NMOS-Gatters absinken kann („Lo“), im Fall der Reihenschaltung höher ist als im Fall der Parallelschaltung. Außerdem nimmt bei der Reihenschaltung der minimal erforderliche Eingangsspannungspegel mit der Anzahl der Schalttransistoren zu, was bei der Parallelschaltung nicht der Fall ist. Dies gilt jedoch nur, wenn die Schalttransistoren in beiden Fällen mit gleichen Kanalweiten und gleichen Kanallängen, d.h. mit gleicher Layout-Fläche ausgelegt werden. Will man gleich niedrige Ausgangspegel erreichen, so müssen die Schalttransistoren der Reihenschaltung dreimal so weit ausgelegt werden wie die der Parallelschaltung, was auf eine dreimal höhere Kapazität pro Eingangsklemme hinausläuft. Deshalb wird in der Praxis die Parallelschaltung meist vorgezogen.

Hinsichtlich der Gleichstrom-Verlustleistung verhalten sich Inverter-, Parallel- und Reihenschaltung gleich. Die höchste Verlustleistung tritt auf, wenn der oder die Schalttransistoren leiten. Dann fällt am Lasttransistor fast die gesamte Betriebsspannung ab, die mit dem fließenden Gleichstrom multipliziert die Verlustleistung ergibt. Wird der Lasttransistor lang und schmal ausgelegt, um diese niedrig zu halten, bedeutet dies andererseits eine lange Aufladezeit der Lastkapazität C_a bei gesperrten Schalttransistoren und damit eine hohe Verzögerungszeit.

8.3.2 Statische PMOS-Logik

Verwendet man anstelle von MOS-Transistoren vom N-Kanaltyp die dazu *komplementären* vom P-Kanaltyp („PMOS-Transistoren“), so werden bei gleichem Schaltungsaufbau die jeweils *komplementären* Schaltfunktionen realisiert. Auch für den Digitalbetrieb von PMOS-Transistoren gelten zwei wohlunterschiedene Spannungspegel, die aus physikalischen Gründen zu denen der NMOS-Logik komplementär sind: mit der Schwellspannung U_{thP} als Trennwert sowie einer unteren Grenze, die u.a. von der Betriebsspannung $U_B' < 0$ und einer oberen Grenze U_o , deren Wert wie im Fall der NMOS-Logik von der Dimensionierung der Schaltung abhängt:

$$(U_B' - U_{thP}) \leq \mathbf{Lo} < U_{thP}$$

$$U_{thP} < \mathbf{Hi} \leq U_o$$

Es gelten folgende Zahlenwerte:

$$U_B' \approx -2,5 \text{ V}$$

$$U_{thP} \approx -0,3 \dots -1,0 \text{ V}$$

Das Bild rechts zeigt eine elementare Inverterschaltung mit PMOS-Transistoren; sie realisiert die logische NICHT-Funktion. Der Inverter besteht aus einem (aktiven) Schalttransistor P_1 und einem zum Zweipol verdrahteten (passiven) Lasttransistor P_L . Man erkennt die Symmetrie zu der in Bild 8.11 gezeigten Inverterschaltung mit NMOS-Transistoren. Da die NICHT-Funktion bekanntlich zu sich selbst komplementär ist, realisieren beide Inverterschaltungen dieselbe Schaltfunktion.

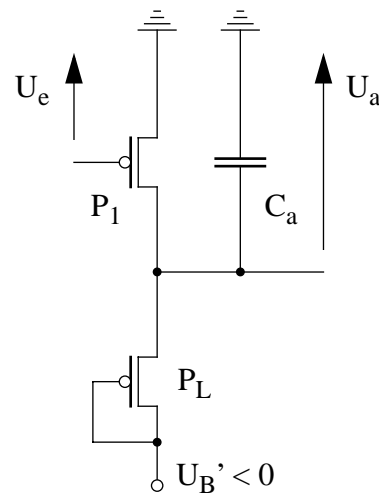


Bild 8.14: Inverterschaltung in statischer PMOS-Logik

Legt man an die Eingangsklemme des PMOS-Inverters einen Potentialpegel, der höher liegt als der (negative) Wert der Schwellspannung des P-Kanals („Hi“), so sperrt der Schalttransistor P_1 . Die Lastkapazität C_a wird dann über den leitenden Lasttransistor P_L geladen („Lo“):

$$\mathbf{Hi:} \quad U_e > U_{thP}$$

$$\mathbf{Lo:} \quad U_a = (U_B' - U_{thP}) < U_{thP}$$

Sobald die Ausgangsspannung auf den Pegel „Lo“ abgesunken ist, erreicht die Gate-Source-Spannung des bisher leitenden Lasttransistors P_L den Wert der Schwellspannung, d.h. er sperrt, so daß sein Sourcepotential U_a nicht weiter absinken kann. Der sich einstellende Ausgangspegel U_a kann direkt als Eingangspegel für gleichartige Schaltungsstufen dienen. Nimmt somit der Eingangspegel den Wert „Lo“ an, liegt er also niedriger als der - negative - Wert der Schwellspannung, so leitet der Schalttransistor P_1 und entlädt (!) die Lastkapazität C_a bis auf den Pegel „Hi“, der durch das Spannungsteilerverhältnis aus dem Schalttransistor P_1 und dem Lasttransistor P_L gegeben ist. Dabei muß P_1 niederohmig sein, d.h. sein Kanalbereich muß breit und kurz ausgelegt werden, P_L dagegen hochohmig, d.h. lang und schmal, da für den Pegel „Hi“ gelten muß:

$$\mathbf{Lo:} \quad U_e = (U_B' - U_{thP}) < U_{thP}$$

$$\mathbf{Hi:} \quad U_a = U_{DS1} > U_{thP}$$

Ersetzt man den Schalttransistor P_1 in der Inverterschaltung (Bild 8.14) durch eine Parallel- oder eine Reihenschaltung mehrerer Schalttransistoren, so erhält man die beiden grundlegenden Verknüpfungsschaltungen in statischer PMOS-Technik, wie sie im folgenden Bild dargestellt sind.

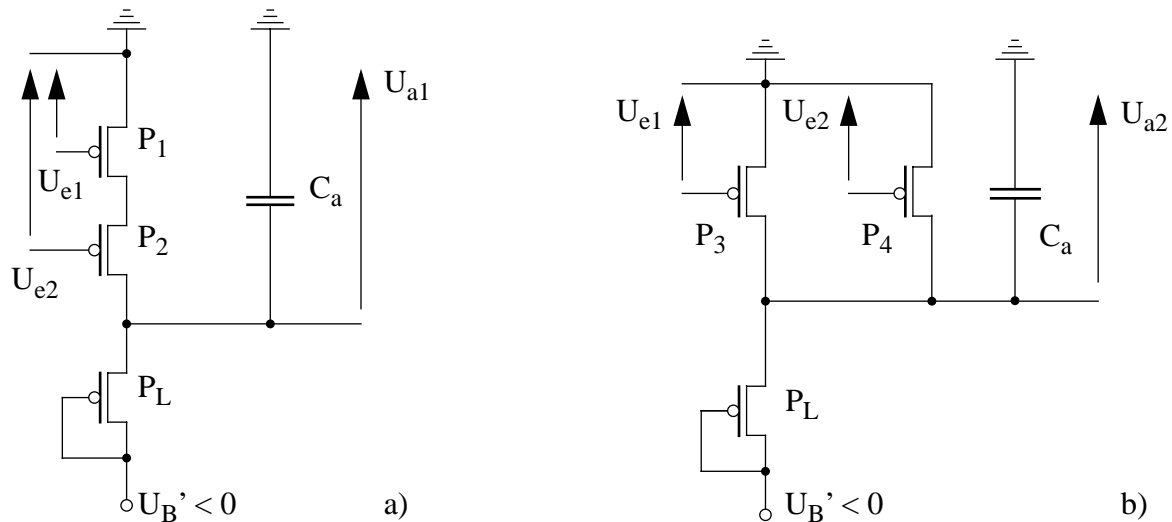


Bild 8.15: Grundsaltungen der statischen PMOS-Logik

U_{e2}	U_{e1}	U_{a1}	U_{e2}	U_{e1}	U_{a2}
Lo	Lo	Hi	Lo	Lo	Hi
Lo	Hi	Lo	Lo	Hi	Hi
Hi	Lo	Lo	Hi	Lo	Hi
Hi	Hi	Lo	Hi	Hi	Lo

Bild 8.16: Realisierung von Schaltfunktionen in statischer PMOS-Logik

a) NOR-Funktion (in positiver Logik) b) NAND-Funktion (in positiver Logik)

Bei einer PMOS-Reihenschaltung nach Bild 8.15a („NOR-Schaltung“) kann die Lastkapazität C_a nur dann entladen werden, wenn alle Schalttransistoren gleichzeitig leiten. Dazu muß das Potential *aller* Eingangsklemmen um mehr als den Betrag der (negativen) Schwellspannung abgesenkt werden. Dabei ist zweierlei zu beachten: Zum einen verändern sich das Source-Potential des Schalttransistors P_2 um den Wert der Drain-Source-Spannungen des darüber liegenden Schalttransistors; zum anderen hängt die Schwellspannung von der Source-Body-Vorspannung und damit vom aktuellen Source-Potential ab. Damit gilt für die anzulegenden Eingangsspannungen und die sich einstellende Ausgangsspannung:

$$\text{Lo: } [U_{e1} < U_{thP1}] \text{ und } [U_{e2} < U_{thP2} + U_{DS1}]$$

$$\text{Hi: } U_{a1} = U_{DS2} + U_{DS1} > U_{thP} \quad (8.24)$$

Wird bei der Reihenschaltung das Potential *mindestens einer* der Eingangsklemmen über den Pegel der (negativen) Schwellspannung angehoben, so sperrt mindestens einer der Schalttransistoren und der hochohmig ausgelegte Lasttransistor P_L lädt die Lastkapazität C_a wie bei der Inverterschaltung

relativ langsam auf den Wert U_{a1} , wobei für PMOS-Transistoren sowohl die Betriebsspannung U_B' als auch die Schwellspannung U_{thP} negativ sind:

$$\mathbf{Hi:} \quad U_{e1} \text{ oder } U_{e2} > U_{thP},$$

$$\mathbf{Lo:} \quad U_{a1} = (U_B' - U_{thP}) < U_{thP} \quad (8.25)$$

In positiver Logik realisiert die PMOS-Reihenschaltung somit eine NOR-, in negativer eine NAND-Funktion, d.h. sie verhält sich wie die NMOS-Parallelschaltung.

Wird bei einer PMOS-Parallelschaltung nach Bild 8.15b („NAND-Schaltung“) das Potential von *mindestens einer* der Eingangsklemmen um mehr als den Betrag der (negativen) Schwellspannung abgesenkt, so wird mindestens einer der Schalttransistoren leitend. Dadurch entlädt sich die Lastkapazität C_a über mindestens einen der niederohmig ausgelegten Schalttransistoren relativ rasch bis auf einen Wert, der durch das Spannungsteilerverhältnis aus den jeweils leitenden Schalttransistoren und dem Lasttransistor P_L gegeben ist, das so dimensioniert werden muß, daß das Ausgangspotential U_{a2} über den Pegel der (negativen) Schwellspannung angehoben wird:

$$\mathbf{Lo:} \quad [U_{e1} < U_{thP}] \text{ oder } [U_{e2} < U_{thP}]$$

$$\mathbf{Hi:} \quad U_{a2} = U_{DS3..4} > U_{thP} \quad (8.26)$$

Wird bei der Parallelschaltung das Potential *aller* Eingangsklemmen über den Pegel der (negativen) Schwellspannung angehoben („Hi“), so sperren alle Schalttransistoren und der hochohmig ausgelegte Lasttransistor P_L lädt die Lastkapazität C_a wie bei der Inverterschaltung relativ langsam auf den Pegel „Lo“:

$$\mathbf{Hi:} \quad U_{e1} \text{ und } U_{e2} > U_{thP}$$

$$\mathbf{Lo:} \quad U_{a2} = (U_B' - U_{thP}) < U_{thP} \quad (8.27)$$

In positiver Logik realisiert die PMOS-Parallelschaltung somit eine NAND-, in negativer eine NOR-Funktion, d.h. sie verhält sich wie die NMOS-Reihenschaltung.

8.3.3 Statische CMOS-Logik

CMOS steht für “Complementary Metal-Oxide-Silicon“, d.h. für die Verwendung von MOS-Transistoren beider, zueinander komplementärer N- bzw. P-Kanaltypen in einer gemeinsamen Schaltung. Auch hierbei handelt es sich um eine statische Logik, obwohl keine Gleichströme fließen, da die zueinander komplementären MOS-Transistoren stets im Wechsel leiten bzw. sperren. Um den Preis eines komplizierteren Fertigungsprozesses kombiniert man NMOS- und PMOS-Transistoren und gewinnt dabei eine Schaltkreisfamilie mit extrem niedriger Verlustleistung.

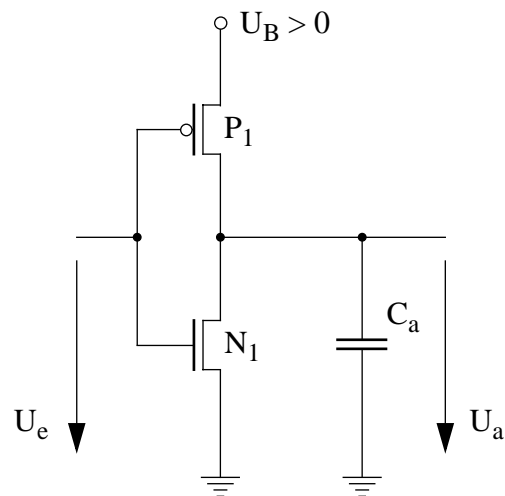


Bild 8.17: Inverterschaltung in statischer CMOS-Logik

Um auch kürzere Umschaltzeiten zu erreichen, kombiniert man die niederohmigen (!) Schalttransistoren des NMOS-Inverters (Bild 8.11) mit denen des PMOS-Inverters (Bild 8.14) und erhält so die im obigen Bild gezeigte elementare Inverterschaltung in statischer CMOS-Technik.

Für den Digitalbetrieb von CMOS-Schaltungen gelten zwei wohlunterschiedene Spannungspegel mit den Schwellspannungen U_{thN} des N-Kanals bzw. U_{thP} des P-Kanals als Trennwerten sowie dem Bezugspotential („Erdpotential“) als unterer und der Betriebsspannung U_B als oberer Grenze, d.h. die Grenzwerte hängen im Gegensatz zur NMOS- oder PMOS-Logik nicht von der Dimensionierung der Bauelemente ab:

$$0 \leq \mathbf{Lo} < [U_{thN} = (U_B + U_{thP})]$$

$$[U_{thN} = (U_B + U_{thP})] < \mathbf{Hi} \leq U_B$$

Es sei an die Zahlenwerte erinnert:

$$U_{thN} \approx +0,3 \text{ .. } +1,0 \text{ V}$$

$$U_{thP} \approx -0,3 \text{ .. } -1,0 \text{ V}$$

$$U_B \approx +2,5 \text{ V}$$

Legt man an die Eingangsklemme des CMOS-Inverters (Bild 8.17) einen Potentialpegel, der niedriger liegt als der Wert der Schwellspannung $U_{thN} > 0$ („Lo“), so sperrt der Schalttransistor N_1 . Gleichzeitig wird der Schalttransistor P_1 leitend, da sein Gatepotential dann niedriger liegt als der (negative) Wert der Schwellspannung $U_{thP} < 0$. Da er niederohmig ausgelegt ist und N_1 sperrt, lädt P_1 die Lastkapazität C_a relativ rasch und vollständig auf den Wert der Betriebsspannung auf („Hi“):

$$\mathbf{Lo:} \quad U_e < [U_{thN} = (U_B + U_{thP})]$$

$$\mathbf{Hi:} \quad U_a = U_B$$

Der sich einstellende Ausgangspegel U_a kann direkt als Eingangspegel für gleichartige Schaltungsstufen dienen. Nimmt somit der Eingangspegel den Wert „Hi“ an, liegt er also höher als die Schwellspannung des N-Kanals, so leitet der ebenfalls niederohmig ausgelegte Schalttransistor N_1 und entlädt die Lastkapazität C_a relativ rasch und vollständig („Lo“), da jetzt der Schalttransistor P_1 sperrt:

$$\mathbf{Hi:} \quad U_e = U_B > [U_{thN} = (U_B + U_{thP})]$$

$$\mathbf{Lo:} \quad U_a = 0$$

Durch Kombination der Schalttransistoren der NOR-Schaltungen für NMOS (Bild 8.12a) und PMOS (Bild 8.15a) erhält man die NOR-Schaltung für die CMOS-Logik im folgenden Bild links, durch Kombination der Schalttransistoren der NAND-Schaltungen für NMOS (Bild 8.12b) und PMOS (Bild 8.15b) die NAND-Schaltung für die CMOS-Logik im nächsten Bild rechts.

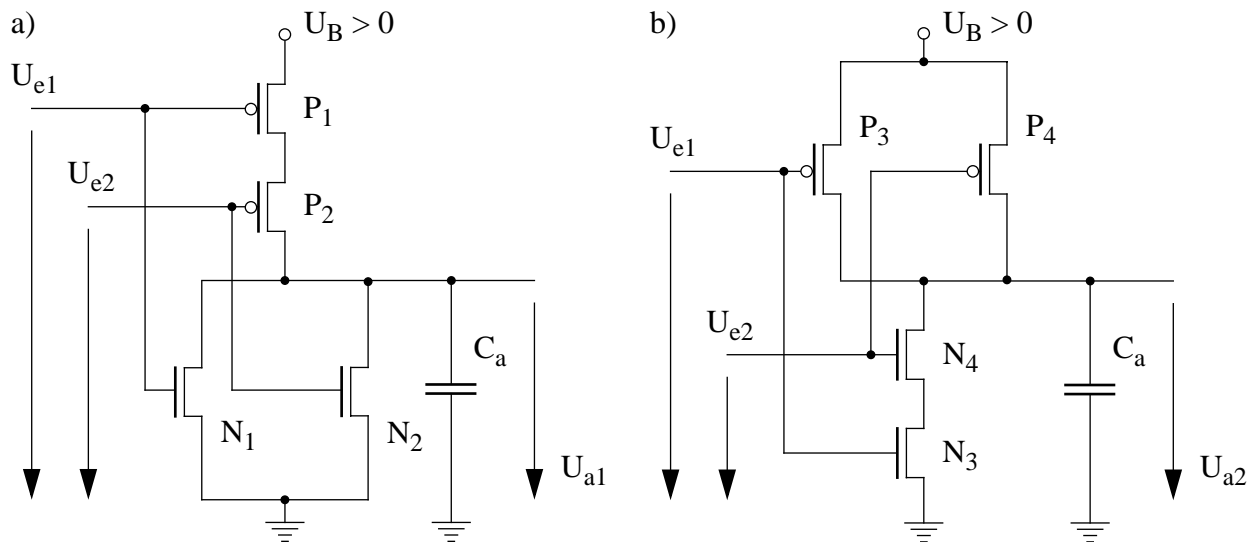


Bild 8.18: Grundsaltungen der statischen CMOS-Logik

U_{e2}	U_{e1}	U_{a1}	U_{e2}	U_{e1}	U_{a2}
Lo	Lo	Hi	Lo	Lo	Hi
Lo	Hi	Lo	Lo	Hi	Hi
Hi	Lo	Lo	Hi	Lo	Hi
Hi	Hi	Lo	Hi	Hi	Lo

Bild 8.19: Realisierung von Schaltfunktionen in statischer CMOS-Logik

a) NOR-Funktion (in positiver Logik)

b) NAND-Funktion (in positiver Logik)

Die in Bild 8.18a gezeigte „NOR-Schaltung“ entspricht im einen Betriebszustand dem CMOS-Inverter nach Bild 8.17 mit leitendem NMOS-Transistor (bei gleichzeitig sperrendem PMOS-Transistor), wenn das Potential von *mindestens einer* der Eingangsklemmen über die Schwellspannung U_{thN} angehoben wird. Dadurch wird die Lastkapazität C_a über mindestens einen der parallel geschalteten, niederohmig ausgelegten NMOS-Transistoren relativ rasch und vollständig entladen, da mindestens einer der in Reihe geschalteten PMOS-Transistoren sperrt:

$$\text{Hi: } U_{e1} \text{ oder } U_{e2} > U_{thN} \qquad \text{Lo: } U_{a1} = 0 < U_{thN} \qquad (8.28)$$

Im anderen Betriebszustand entspricht die Schaltung dem Inverter mit sperrendem NMOS-Transistor (bei gleichzeitig leitendem PMOS-Transistor), wenn das Potential *aller* Eingangsklemmen unter die Schwellspannung U_{thN} abgesenkt wird. Dadurch wird die Reihenschaltung der niederohmig ausgelegten PMOS-Transistoren leitend und lädt die Lastkapazität C_a relativ rasch und voll-

ständig auf den Wert der Betriebsspannung auf, da alle parallel geschalteten NMOS-Transistoren sperren:

$$\text{Lo: } U_{e1} \text{ und } U_{e2} < U_{thN} \qquad \text{Hi: } U_{a1} = U_B > U_{thN} \qquad (8.29)$$

Somit realisiert die in Bild 8.18a gezeigte CMOS-Schaltung in positiver Logik eine NOR-, in negativer eine NAND-Funktion.

Die in Bild 8.18b gezeigte „NAND-Schaltung“ entspricht im einen Betriebszustand dem CMOS-Inverter nach Bild 8.17 mit leitendem NMOS-Transistor (bei gleichzeitig sperrendem PMOS-Transistor), wenn das Potential *aller* Eingangsklemmen über die Schwellspannung U_{thN} angehoben wird. Dadurch wird die Reihenschaltung der niederohmig ausgelegten NMOS-Transistoren leitend und entlädt die Lastkapazität C_a relativ rasch und vollständig, da alle parallel geschalteten PMOS-Transistoren sperren:

$$\text{Hi: } U_{e1} > \text{ und } U_{e2} > U_{thN} \qquad \text{Lo: } U_{a2} = 0 < U_{thN} \qquad (8.30)$$

Im anderen Betriebszustand entspricht die Schaltung dem Inverter mit sperrendem NMOS-Transistor (bei gleichzeitig leitendem PMOS-Transistor), wenn das Potential von *mindestens einer* der Eingangsklemmen unter die Schwellspannung U_{thN} abgesenkt wird. Dadurch wird die Lastkapazität C_a über mindestens einen der parallel geschalteten, niederohmig ausgelegten PMOS-Transistoren relativ rasch und vollständig auf den Wert der Betriebsspannung aufgeladen, da mindestens einer der in Reihe geschalteten NMOS-Transistoren sperrt:

$$\text{Lo: } U_{e1} \text{ oder } U_{e2} < U_{thN} \qquad \text{Hi: } U_{a2} = U_B > U_{thN} \qquad (8.31)$$

Somit realisiert die in Bild 8.18b gezeigte CMOS-Schaltung in positiver Logik eine NAND-, in negativer eine NOR-Funktion.

Fassen wir die Vorteile der CMOS-Schaltkreisfamilie zusammen:

- Niedrige Verlustleistung, da über die zueinander komplementären, im Gegentakt leitenden bzw. sperrenden MOS-Transistoren keine Gleichströme fließen. Beim Umschalten wird nur Wechselstromverlustleistung umgesetzt.
- Hohe Schaltgeschwindigkeit durch die Kombination ausschließlich niederohmiger Schalttransistoren, d.h. dank der Vermeidung hochohmiger Lastwiderstände.
- Hoher Störabstand durch Aufladen der zu treibenden Lastkapazität auf den vollen Wert der Betriebsspannung bzw. deren vollständiges Entladen dank der Vermeidung eines Spannungsteilers.

8.3.4 Dynamische Domino-Logik

Wegen der inhärenten Gatekapazitäten von MOS-Transistoren können die damit aufgebauten Schaltungen auch *dynamisch* betrieben werden. Sie entwickeln dann keine statische Verlustleistung, da im Betrieb keine Gleichströme fließen. Die hier vorgestellte Domino-Schaltungstechnik wird in CMOS-Technologie realisiert, d.h. sie verwendet MOS-Transistoren vom N-Kanal- und vom P-Kanaltyp. Deshalb können die Schaltungen mit einem einzigen Taktsignal betrieben werden.

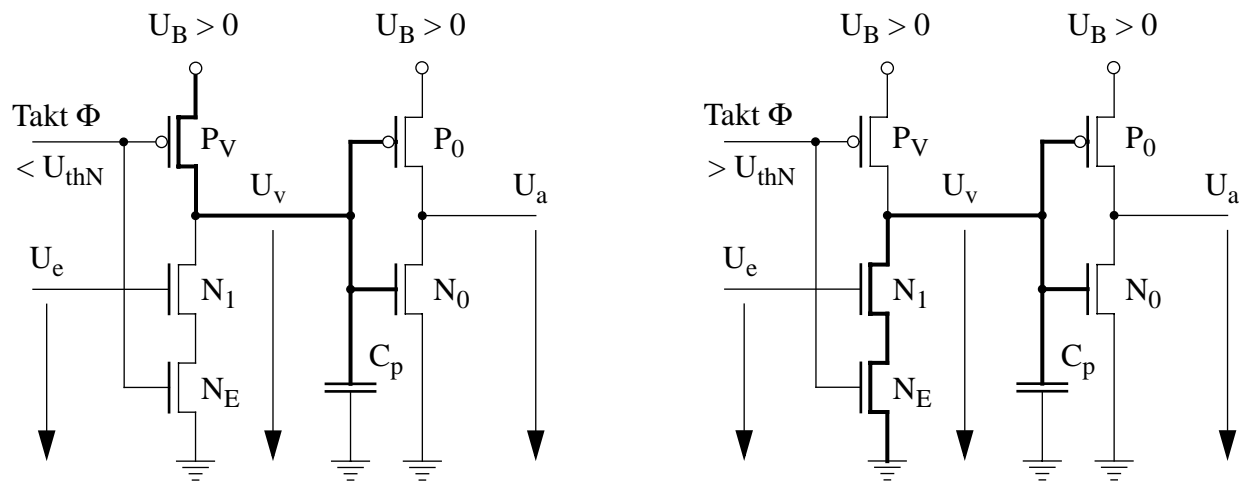


Bild 8.20: Grundschaltung in CMOS-Domino-Logik

a) unbedingte Vorladephase

b) bedingte Evaluierungsphase

Den Aufbau einer Grundschaltung in Domino-Logik zeigt das obige Bild. Ein PMOS-Transistor P_V und ein NMOS-Transistor N_E werden gemeinsam von einem Taktsignal Φ angesteuert; sie leiten/sperren jeweils wechselseitig, so daß kein Gleichstrom fließt. Zu diesem, im Gegentakt arbeitenden CMOS-Transistorpaar ist der von einem Eingangssignal U_e angesteuerte Schalttransistor N_1 in Reihe geschaltet. Die Ausgangsstufe ist ein CMOS-Inverter, dessen inhärente Gatekapazität C_p für den dynamischen Betrieb der Schaltung verwendet wird, wobei zwei Phasen zu unterscheiden sind.

- Während der ersten, der Vorladephase gilt:

$$\text{Lo: Taktpegel } \Phi < [U_{thN} = U_B + U_{thP}] \quad (8.32)$$

Der Vorladetransistor P_V leitet und lädt die Kapazität C_p , da der Evaluierungstransistor N_E sperrt, unabhängig vom Leitungszustand des Schalttransistors N_1 auf den Wert der Betriebsspannung $U_B \approx 2,5 \text{ V}$ auf (Bild 8.20a). Deshalb leitet der Transistor N_0 des Ausgangsinverters und senkt den Ausgangsspannungspegel („unbedingt“) auf Erdpotential ab:

$$U_e \text{ beliebig} \quad \text{Lo: } U_a = 0 < U_{thN} \quad (8.33)$$

- Während der zweiten, der Evaluierungsphase gilt:

$$\text{Hi: Taktpegel } \Phi > [U_{thN} = U_B + U_{thP}] \quad (8.34)$$

Der Evaluierungstransistor N_E leitet und entlädt, falls abhängig vom Eingangssignal U_e der Schalttransistor N_1 ebenfalls leitet, die Kapazität C_p , da der Vorladetransistor P_V sperrt (Bild 8.20b).

Dann leitet der Transistor P_0 des Ausgangsinverters und hebt den Ausgangsspannungspegel („bedingt“) auf den Wert der Betriebsspannung U_B an:

$$\text{Hi: } U_e = U_B > U_{thN}$$

$$\text{Hi: } U_a = U_B > U_{thN} \quad (8.35)$$

Andernfalls, wenn der Schalttransistor N_1 sperrt, kann der Evaluierungstransistor N_E die Kapazität C_p nicht entladen. Dann bleibt der Transistor N_0 des Ausgangsinverters leitend und hält den Ausgangsspannungspegel („bedingt“) auf Erdpotential:

$$\text{Lo: } U_e = 0 < U_{thN}$$

$$\text{Lo: } U_a = 0 < U_{thN} \quad (8.36)$$

Man erkennt, daß die Grundsaltung in Domino-Logik die Schaltfunktion der „Identität“ realisiert. Sie ist, wie auch eine logische Verknüpfung mehrerer Eingangssignale, nur während der Evaluierungsphase gültig.

Schaltet man nämlich zwischen das vom Taktsignal Φ angesteuerte CMOS-Transistorpaar eine Reihen- oder eine Parallelschaltung von NMOS-Schalttransistoren (oder bei komplizierteren Schaltfunktionen auch direkt ein komplexeres Netzwerk aus NMOS-Transistoren), so erhält man in positiver Logik eine UND-Schaltung wie im nächsten Bild links, bzw. eine ODER-Schaltung wie im nächsten Bild rechts dargestellt. Aufgrund des Ausgangsinverters, der die für den dynamischen Betrieb erforderliche Kapazität beisteuert und außerdem eine Signalverstärkung liefert, sind die in Domino-Logik realisierten Schaltfunktionen grundsätzlich bejaht. Schaltet man mehrere Domino-Stufen in einer Kette hintereinander, so setzen sich die Verknüpfungen der Eingangssignale von Stufe zu Stufe fort (daher der Name dieser Schaltungstechnik).

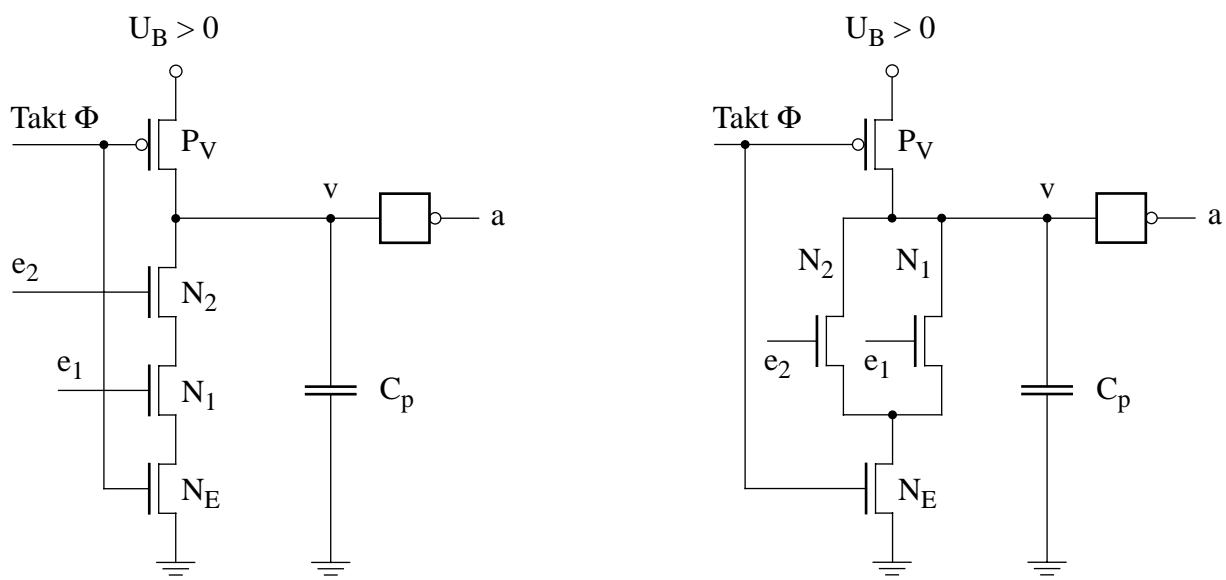


Bild 8.21: CMOS-Domino-Logik zur Realisierung von Schaltfunktionen

a) UND-Schaltung (in positiver Logik) b) ODER-Schaltung (in positiver Logik)

Literatur: R. H. Krambeck, C. M. Lee, H. F. S. Law, “High Speed Compact Circuits With CMOS“, IEEE Journal of Solid-State Circuits, Vol. SC-17, No. 3, pp. 614-619 (June 1982).

9.1 Speicherzellen mit MOS-Transistoren

Aufgrund seiner Bau- und Wirkungsweise ist jeder MOS-Transistor mit einer Gate-Kapazität behaftet, die es erlaubt, sowohl *statisch* als auch *dynamisch* betriebene MOS-Speicherzellen herzustellen. Die statischen Zellen beruhen auf dem Prinzip des bistabilen Flipflops, die dynamischen auf der Ladungsspeicherung in Kondensatoren. Dynamische MOS-Speicher (DRAM) haben einen niedrigeren Flächenbedarf pro Speicherzelle und eine niedrigere Verlustleistung, so daß sich größere Speicher auf einem Chip integrieren lassen. Statische MOS-Speicher (SRAM) haben eine geringere Zugriffszeit, da sie (im Gegensatz zu den dynamischen) keine Auffrischzyklen benötigen.

Zur Einführung in die Schaltungstechnik zeigt das folgende Bild die Implementierung eines ungetakteten Basis- und eines zustandsgetakteten Auffang-Flipflops mit logischen Gattern. Die Realisierung kann grundsätzlich in allen Schaltungstechnologien erfolgen; hier wird die Realisierung in statischer NMOS-Logik gezeigt.

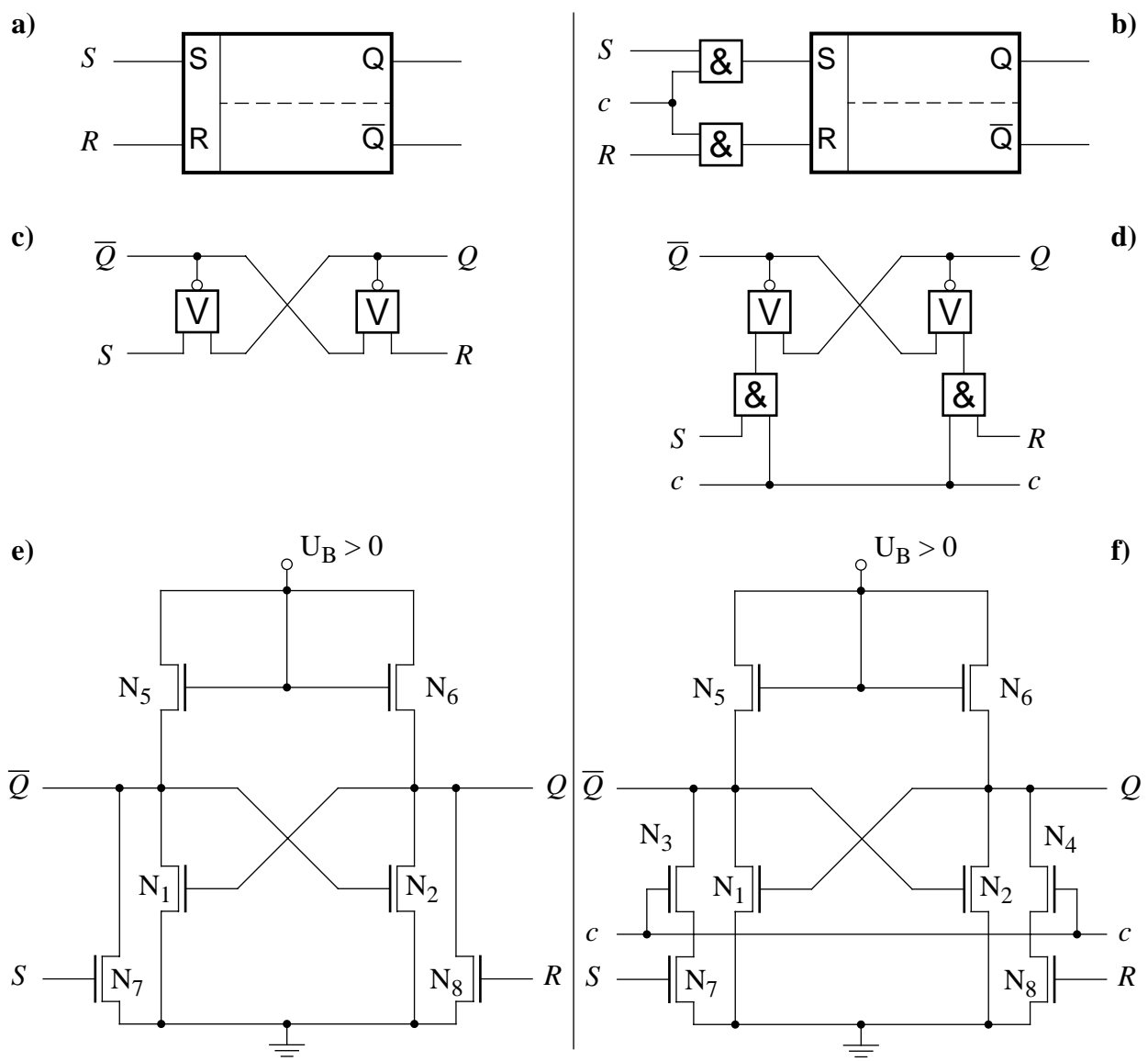


Bild 9.1: Basis-Flipflop (links) und Auffang-Flipflop (rechts). a) b) Schaltsymbol; c) d) Implementierung mit logischen Gattern; e) f) Realisierung mit NMOS-Transistoren.

9.1.1 Statische Speicherzelle

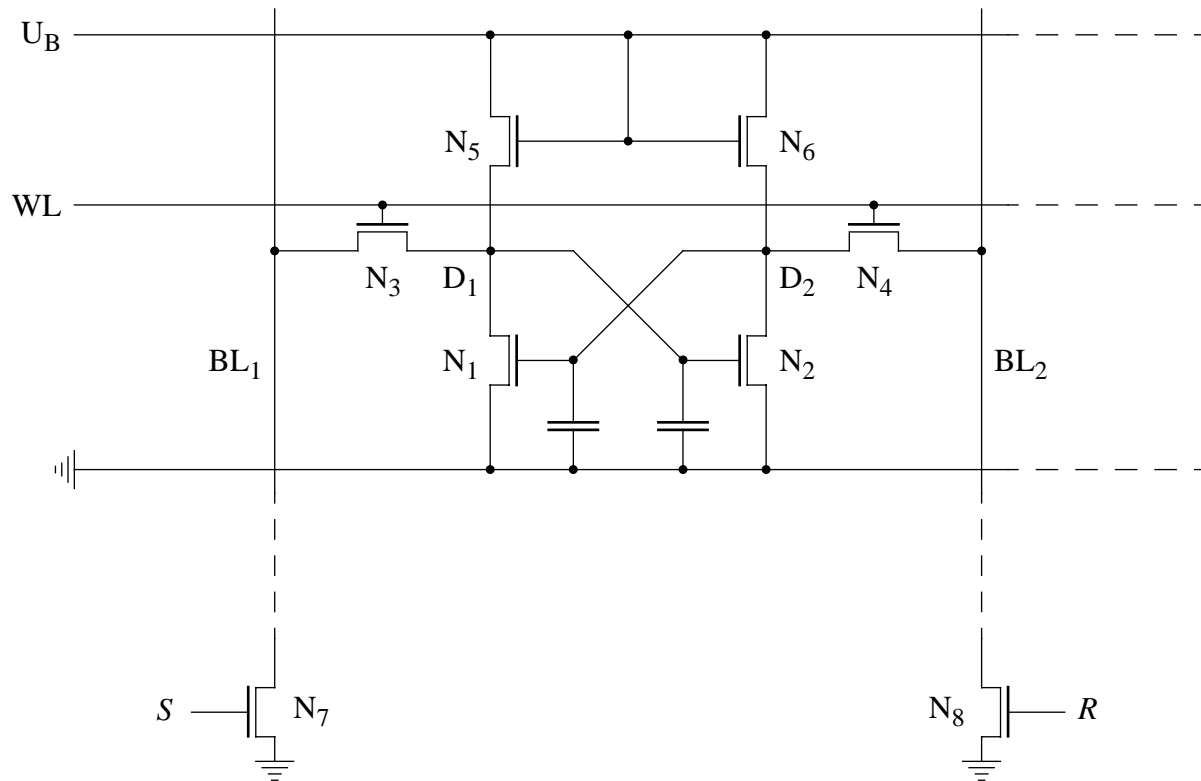


Bild 9.2: Schaltbild der statischen „Sechs-Transistor-Speicherzelle“ mit Lasttransistoren N_5 , N_6 .
WL - Wortleitung; BL_1 , BL_2 - Bitleitungspaar. - Literatur: J. S. Schmidt, „Integrated MOS
Random-access Memory“, Solid-State Design, 21-25 (1965).

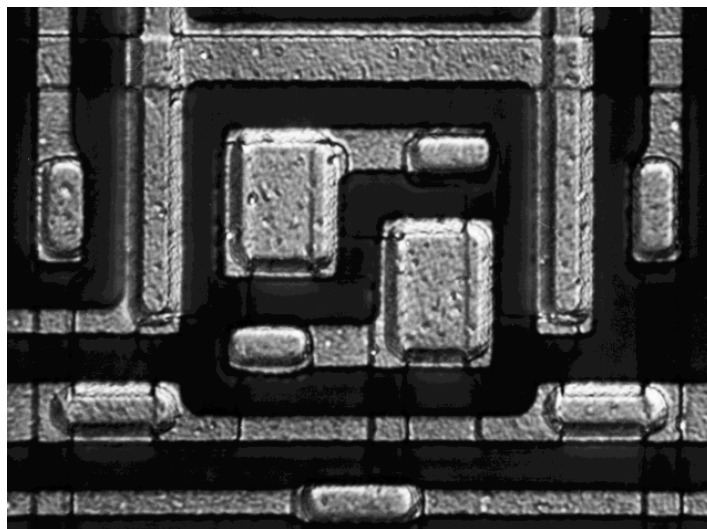


Bild 9.3: Layout der statischen „Sechs-Transistor-Speicherzelle“.
Literatur: R. Remshardt, U.G. Baitinger: „A High Performance Low Power 2048-bit
Memory Chip in MOSFET Technology And Its Application“, IEEE J. Solid-State Cir-
cuits, Vol. SC-11, No.3 (June 1976), pp. 352-359

Die beiden obigen Bilder zeigen das Schaltbild und das Layout einer bistabilen Speicherzelle aus sechs NMOS-Transistoren ("Six-Device Cell"): es handelt sich um MOS-Transistoren mit N-Kanal vom Anreicherungstyp. Anhand dieses Transistortyps läßt sich die Wirkungsweise einer Speicherzelle anhand positiver Spannungswerte in positiver Logik erläutern, was das Verständnis erleichtert. Die Speicherzelle kann mit einer Betriebsspannung von $U_B \approx 2,5 \text{ V}$ betrieben werden.

a) Betriebsart „Speichern“

Wenn die Speicherzelle unselektiert ist, liegt das Potential der Wortleitung WL unterhalb der Schwellspannung U_{thN} („threshold voltage“); dann sind die Ein/Ausgangstransistoren N_3 und N_4 gesperrt:

$$\text{Lo: } U_{WL} \approx 0 < U_{thN} \approx 0,3 \dots 1,0 \text{ V}$$

Leitet Schalttransistor N_1 , was dem gespeicherten Zustandsbit $Q = 1$ entsprechen möge, so liegen der Drainknoten D_1 und damit das Gate des Schalttransistors N_2 auf dem Potential „Lo“, d.h. unterhalb der Schwellspannung, da der Schalttransistor N_1 niederohmig gegenüber dem Lasttransistor N_5 ausgelegt ist. Dadurch wird Schalttransistor N_2 gesperrt. Das hat zur Folge, daß der Drainknoten D_2 über den Lasttransistor N_6 auf das Potential „Hi“ angehoben wird.

$$\text{Lo: } U_{D1} < U_{thN2}$$

$$\text{Hi: } U_{D2} = (U_B - U_{thN6}) > U_{thN1}$$

Dadurch bleibt Schalttransistor N_1 leitend, wie ursprünglich angenommen. - In gleicher Weise kann sich das komplementäre Zustandsbit $Q = 0$ stabil halten, da die Schaltung symmetrisch ist.

b) Betriebsart „Lesen“

Die Speicherzelle wird selektiert, indem das Potential der Wortleitung WL auf den Wert der Betriebsspannung U_B angehoben wird:

$$\text{Hi: } U_{WL} = U_B > U_{thN}$$

Dadurch werden die Ein/Ausgabetransistoren N_3 und N_4 leitend und verbinden die Zellknoten D_1 und D_2 mit den Bitleitungen BL_1 bzw. BL_2 , die beide auf mindestens den Wert der Betriebsspannung aufgeladen sein mögen:

$$\text{Hi: } U_{BL1} \geq U_B$$

$$\text{Hi: } U_{BL2} \geq U_B$$

Leitet Schalttransistor N_1 , während Schalttransistor N_2 sperrt, was dem gespeicherten Zustandsbit $Q = 1$ entsprechen möge, so fließt über den relativ hochohmigen Ein/Ausgabetransistor N_3 in Reihe mit dem niederohmigen Schalttransistor N_1 ein Strom, so daß am Zellknoten D_1 der Potentialpegel „Lo“ erhalten bleibt, während der gesättigt leitende Ein/Ausgabetransistor N_4 den Zellknoten D_2 auf dem Potentialpegel „Hi“ zumindest hält.

$$\text{Lo: } U_{D1} < U_{thN2}$$

$$\text{Hi: } U_{D2} = (U_B - U_{thN4}) > U_{thN1}$$

Wesentlich ist, daß das Gatepotential U_{D2} des als leitend angenommenen Schalttransistors N_1 sogar ansteigt, so daß sich über den Ein/Ausgabetransistor N_3 ein kräftiger Lesestrom ergibt, der aus Bitleitung BL_1 nachgeliefert wird. Er kann mit einem selektiv an das Bitleitungspaar angeschlossenen stromempfindlichen Differenzverstärker abgefühlt werden. - Ist in der Zelle das komplementäre Zustandsbit $Q = 0$ gespeichert, fließt der Lesestrom in symmetrischer Weise aus Bitleitung BL_2 .

c) Betriebsart „Schreiben“

Soll in die Speicherzelle geschrieben, d.h. das Zustandsbit $Q \in \{0, 1\}$ geändert werden, wird sie wie beim Lesen durch Anheben des Potentials der Wortleitung WL selektiert, so daß die Ein/Ausgabetransistoren N_3 und N_4 leiten:

$$\text{Hi: } U_{WL} = U_B > U_{thN}$$

Soll der Schalttransistor N_2 leitend werden, was wie oben angenommen dem gespeicherten Zustandsbit $Q = 0$ entspricht, so muß die Bitleitung BL_2 auf Erdpotential, d.h. unter den Wert der Schwellspannung abgesenkt werden, während die Bitleitung BL_1 mindestens auf dem Wert der Betriebsspannung bleibt:

$$\text{Hi: } U_{BL1} \geq U_B$$

$$\text{Lo: } U_{BL2} = 0 < U_{thN}$$

Der gesättigt leitende Ein/Ausgabetransistor N_3 hebt das Potential des Zellknotens D_1 und damit das Gatepotential von Schalttransistor N_2 über den Wert der Schwellspannung an, so daß dieser leitet, während Schalttransistor N_1 gesperrt wird, da der Zellknoten D_2 über den linear leitenden Ein/Ausgabetransistor N_4 das niedrige Potential der angeschlossenen Bitleitung BL_2 annimmt :

$$\text{Hi: } U_{D1} = (U_B - U_{thN3}) > U_{thN2}$$

$$\text{Lo: } U_{D2} = U_{BL2} = 0 < U_{thN1}$$

Man erkennt, daß die *Geschwindigkeit*, mit der die statische Speicherzelle beim Schreiben ihren Zustand ändert, offenbar von der Aufladung der Gatekapazität des jeweils einzuschaltenden Schalttransistors über einen Ein/Ausgabetransistor abhängt. Das entspricht ganz der Wirkungsweise einer Inverterschaltung. Mit anderen Worten: Diese Speicherzelle wird von zwei kreuzgekoppelten Invertern gebildet, die aus Schalttransistor N_1 und Ein/Ausgabetransistor N_3 bzw. Schalttransistor N_2 und Ein/Ausgabetransistor N_4 bestehen. - Die Aufgabe der Lasttransistoren N_5 und N_6 besteht lediglich darin, während des Ruhezustands der Speicherzelle die Leckströme an den Drainknoten D_1 bzw. D_2 nachzuliefern (die dort zum Substrat abfließen), damit das Gatepotential des jeweils leitenden Schalttransistors nicht unter die Schwellspannung absinkt, so daß er leitend bleibt.

Die *Verlustleistung* dieser Speicherzelle wird hauptsächlich in dem Lasttransistor N_5 oder N_6 in Wärme umgesetzt, der sich in Reihe zum jeweils leitenden Schalttransistor befindet, da an ihm fast die gesamte Betriebsspannung ($U_B - U_{D1} \approx U_B$) abfällt (leider nutzlos, da er keinen Leckstrom nachzuliefern hat, weil er zur entladenen Gatekapazität führt). - Eine elegante Methode, dies zu vermeiden, besteht darin, die Lasttransistoren mit dem komplementären P-Kanaltyp auszuführen. Dadurch wird gewährleistet, daß Lasttransistor N_5 sperrt, wenn Schalttransistor N_1 leitet, so daß praktisch keine Gleichstromverlustleistung verbraucht wird, während andererseits Lasttransistor N_6 leitet, um den Leckstrom an der aufgeladenen Gatekapazität des leitenden Schalttransistors N_1 zu kompensieren. Man erkaufte den Vorteil einer extrem niedrigen Verlustleistung der integrierten Speichermatrizen durch den komplizierteren CMOS-Herstellungsprozeß.

9.1.2 Dynamische Speicherzellen

a) Vier-Transistor-Speicherzelle

Wir haben im letzten Abschnitt gesehen, daß bei der Leseoperation einer statischen Speicherzelle die Gatekapazität des leitenden Schalttransistors über einen Ein/Ausgabetransistor auf die Spannung der angeschlossenen Bitleitung aufgeladen wird. Die Aufgabe eines Lasttransistors besteht nur darin, den Leckstrom dieser Kapazität während des Ruhezustands zu kompensieren, damit die Gate-Source-Spannung des leitenden Schalttransistors nicht unter den Wert seiner Schwellspannung U_{th} absinkt. Ist der erwähnte Leckstrom klein genug, so sinkt die Spannung an der aufgeladenen Gatekapazität so langsam ab, daß sich der aktuelle Zustand der Speicherzelle während mehrerer Lesezyklen (anderer Speicherzellen in der Speichermatrix) halten kann. Wird rechtzeitig eine Leseoperation an der betreffenden Speicherzelle ausgeführt, so wird der Ladungszustand der Gatekapazität und damit das gespeicherte Bit *aufgefrischt*. Dann können die relativ großen Lasttransistoren entfallen und man erhält die aus vier NMOS-Transistoren bestehende Speicherzelle ("Four-Device Cell"), die im folgenden Bild gezeigt wird.

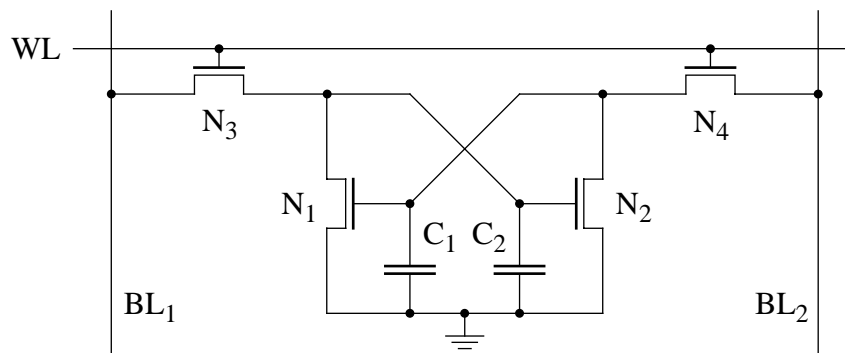


Bild 9.4: Dynamische „Vier-Transistor-Speicherzelle“ mit zwei Ladungsspeichern (C_1 , C_2).

WL - Wortleitung; BL_1 , BL_2 - Bitleitungspaar.

Literatur: B. R. Tunzi, "MOS Random Access Arrays", Electronics, 103 (1969).

Obwohl die obige Schaltung aufgrund ihrer Symmetrie einem Flipflop ähnelt, ist hier das gespeicherte Bit nicht im (gegensätzlichen) Leitungszustand der Schalttransistoren N_1 und N_2 gespeichert, sondern als elektrische Ladung in einer ihrer Gate-Kapazitäten C_1 oder C_2 .

- Der Ladungszustand der MOS-Kapazitäten (C_1 , C_2 im obigen Bild) bedarf im Ruhezustand der Speicherzelle einer regelmäßigen Auffrischung, was als *dynamischer Betrieb* bezeichnet wird.

Die Auffrischung kann zweckmäßig so erfolgen, daß die Wortleitung WL regelmäßig selektiert wird, wobei alle daran angeschlossenen Speicherzellen in derselben Zeile der Speichermatrix aufgefrischt werden, wobei ein Anschließen des Leseverstärkers nicht erforderlich ist. Dadurch wird die *Zugriffszeit* der dynamischen 4-Transistor-Speicherzelle im Durchschnitt zwar größer als die der statischen 6-Transistor-Speicherzelle (Bild 9.2), da sie während der Auffrischzyklen weder gelesen noch geschrieben werden kann. Doch kann durch Einsparen der Lasttransistoren die Zellfläche stark verkleinert und die *Verlustleistung* verringert werden, so daß sich Matrizen größerer Speicherkapazität auf derselben Chipfläche integrieren lassen.

b) Drei-Transistor-Speicherzelle

Während die dynamische Vier-Transistor-Speicherzelle (Bild 9.4) symmetrisch aufgebaut ist und wie eine statische gelesen und geschrieben werden kann, erfordern die unsymmetrischen (dynamischen) Speicherzellen, die nur *einen* Ladungsspeicher enthalten, andere Betriebsarten. Ein Beispiel einer solchen Speicherzelle zeigt das folgende Bild. Die Schaltung besteht aus einem Speicherkondensator C_1 und drei NMOS-Transistoren ("Three-Device Cell").

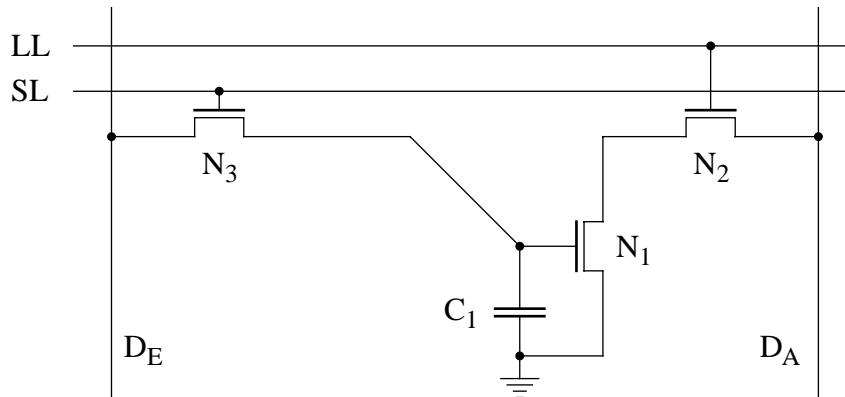


Bild 9.5: Dynamische „Drei-Transistor-Speicherzelle“ mit einem Ladungsspeicher (C_1).

LL - Leseleitung; SL - Schreibleitung; D_E - Dateneingangsleitung;

D_A - Datenausgangsleitung. - Literatur: W. M. Regitz, J. Karp, "A Three Transistor 1024 bit 500 ns MOS RAM", IEEE J. Solid-State Circuits 5, 182-186 (1970).

Das gespeicherte Bit entspricht dem geladenen bzw. ungeladenen Zustand des Kondensators C_1 . Die Zuordnung des aktuellen Werts des Bits zum Ladungszustand kann frei definiert werden.

Ist der Gate-Kondensator C_1 geladen, so liegt das Gatepotential des Schalttransistors N_1 höher als seine Schwellspannung U_{th} , so daß er leitet. Um das gespeicherte Bit zu lesen, wird das Potential der selektierenden Leseleitung LL über die Schwellspannung angehoben, damit auch der Ausgabetransistor N_2 leitet. (Der Eingabetransistor N_3 bleibt gesperrt.) Nun kann sich die zuvor aufgeladene Datenausgangsleitung D_A über den Ausgabetransistor N_2 und den in Reihe dazu liegenden Schalttransistor N_1 auf Erdpotential entladen. Der Entladestrom wird von einem an die Datenausgangsleitung D_A angeschlossenen Leseverstärker festgestellt. - Ist bei Speicherung des kompletären Bits der Kondensator C_1 entladen, so sperrt Schalttransistor N_1 und die aufgeladene Datenausgangsleitung D_A entlädt sich nicht.

Um ein Bit einzuschreiben, wird das Potential der Schreibleitung SL über die Schwellspannung angehoben, so daß der Eingabetransistor N_3 leitet (während der Ausgabetransistor N_2 sperrt). Je nach Potential der Dateneingangsleitung D_E wird der Kondensator C_1 aufgeladen oder entladen.

Das unbedingt erforderliche regelmäßige Auffrischen des Ladezustands der Speicherzelle ist relativ umständlich. Zunächst muß das gespeicherte Bit einer aufzufrischenden Speicherzelle gelesen und an die zugehörige Dateneingangsleitung D_E angelegt werden. Wird mit allen Speicherzellen innerhalb einer Zeile so verfahren, befinden sich alle Dateneingangsleitungen, d.h. alle Spalten der Speichermatrix auf entsprechenden Potentialen. Dann darf nur die Schreibleitung SL selektiert werden, die der genannten Zeile zugehört. Damit werden die auf- bzw. entladenen Dateneingangs-

der relativ groß ist im Vergleich zum Kapazitätswert des Speicherkondensators C. Es ist unbedingt erforderlich, eine gelesene Zelle unmittelbar danach aufzufrischen.

Der Vorteil der dynamischen Ein-Transistor-Speicherzelle liegt in ihrer minimalen Zellfläche; er muß jedoch durch entsprechend aufwendige Peripherieschaltkreise erkauft werden. Die niedrige Verlustleistung erlaubt die Integration extrem großer Speichermatrizen auf gegebener Chipfläche. Wegen des erforderlichen komplizierten Auffrischvorgangs ist die Zugriffszeit relativ hoch.

9.2 Speichermatrizen (RAM)

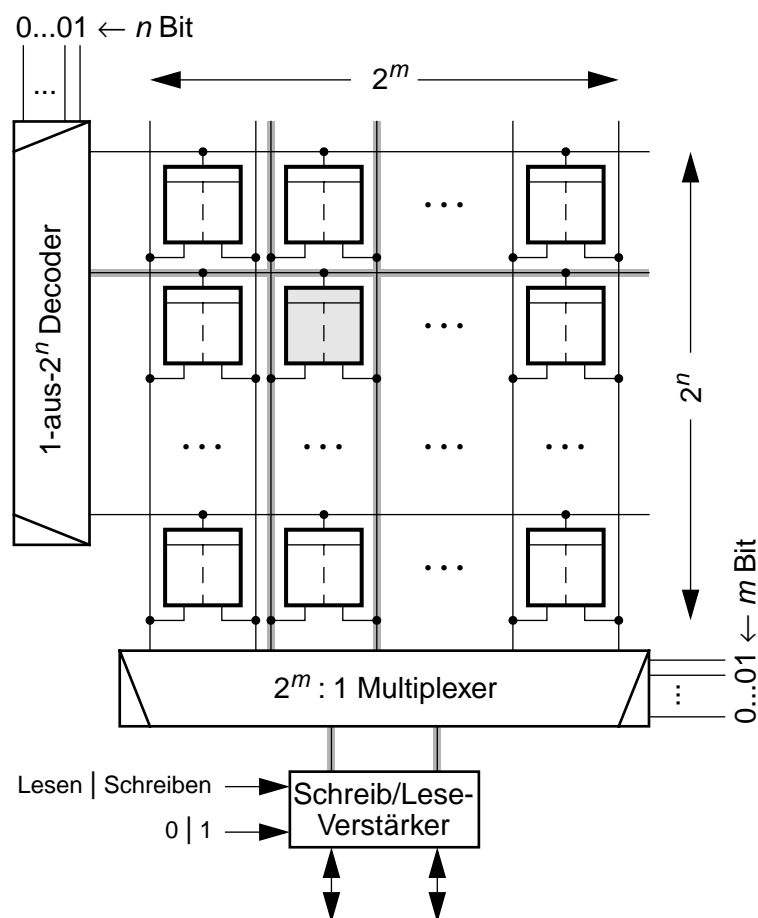


Bild 9.8: Blockschaltbild eines Schreib/Lese-Speichers mit wahlfreiem Zugriff (RAM)

Die Technologie der Mikroelektronik erlaubt es, Speicherzellen in extrem großer Anzahl auf einem gemeinsamen Silizium-Chip zu integrieren: Stand der Technik sind mehrere Millionen Bits pro Chip. Die Speicherzellen sind geometrisch in Zeilen und Spalten angeordnet. Eine Zeile von Speicherzellen kann z.B. ein Wort enthalten, dessen einzelne Bits sich in unterschiedlichen Spalten befinden. Man bezeichnet daher die Zeilen der Speichermatrix auch als „Wortrichtung“, die Spalten als „Bitrichtung“. Selektiert man gleichzeitig eine Zeile und eine Spalte der Speichermatrix, so erhält man den direkten Zugriff auf die Speicherzelle am Schnittpunkt. Dies wird als wahlfreier Zugriff bezeichnet („Random Access Memory“, RAM).

Die Selektion einer Speicherzelle auf einem höchstintegrierten Speicherchip kann nicht direkt erfolgen. Eine Speichermatrix der Größe

$$2^n \text{ Zeilen} \cdot 2^m \text{ Spalten} = 2^{n+m}$$

$$\text{Beispiel: } 2^{10+10} = 2^{20} = 1 \text{ MegaBit,}$$

die also etwa einer Million Speicherzellen enthält, würde sonst

$$2^n \text{ Zeilen} + 2^m \text{ Spalten} = 2^n + 2^m$$

$$2^{10} + 2^{10} = 1.024 + 1.024 = \underline{2.048} (!)$$

äußere Anschlüsse benötigen, was technologisch nicht machbar ist. Codiert man jedoch die Zeilen- und Spaltenadressen („Wort- bzw. Bitadresse“) im Dualcode, so gilt für die Anzahl der äußeren Anschlüsse:

$$ld\ 2^n + ld\ 2^m = n + m$$

$$10 + 10 = \underline{20} .$$

Zur Selektion genau einer Wort- bzw. einer Bitleitung müssen die dual codierten Wort- bzw. Bitadressen in einen „1-aus- n “-Code umgewandelt werden, bei dem definitionsgemäß immer nur ein Bit aktiv ist. Daher enthalten höchstintegrierte Speicherchips am Rande der eigentlichen Speichermatrix stets auch Decoderschaltungen, um die Anzahl der äußeren Anschlüsse des Chips in der geschilderten Weise niedrig zu halten, wie das vorhergehende Bild zeigt. Das folgende Bild zeigt das Layout einer Speichermatrix mit wahlfreiem Zugriff (RAM).

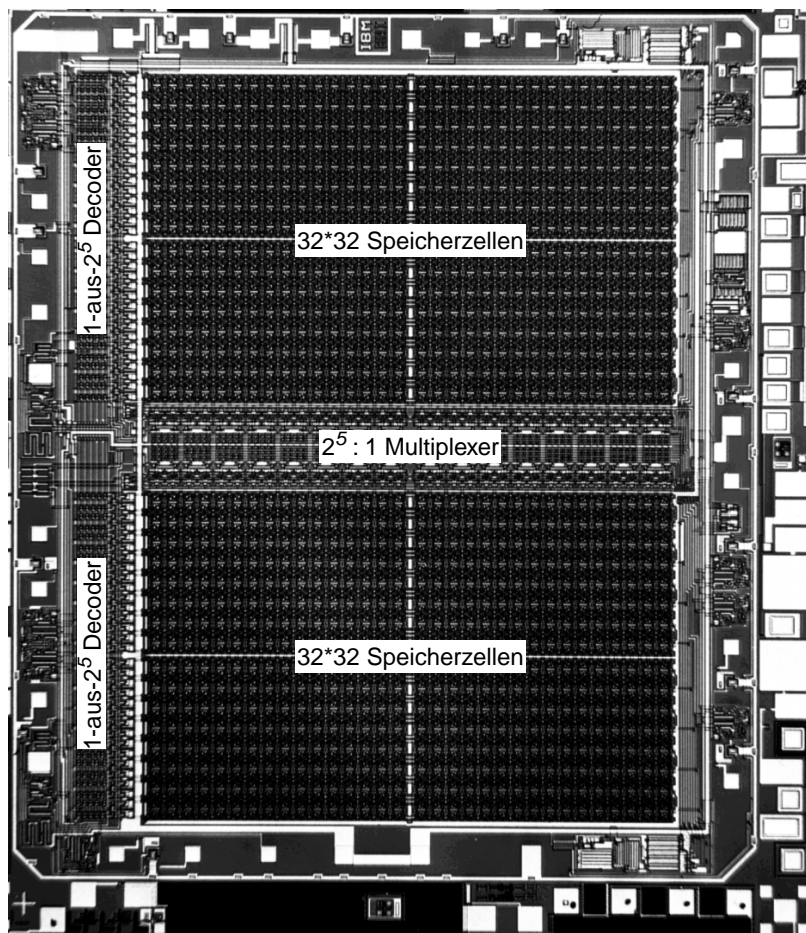


Bild 9.9: Chip-Layout eines Schreib/Lese-Speichers mit wahlfreiem Zugriff (32*64 bit RAM)

10.1 Technologie und Schaltungstechnik

10.1.1 Technologische Randbedingungen

Die *Mikroelektronik* ermöglicht die Integration umfangreicher und komplexer elektronischer Schaltungen auf einem einzigen Halbleiterchip. Insbesondere die Siliziumtechnologie erlaubt es, dank der Eigenschaften des Siliziumdioxids, kleinste Strukturen im Mikrometerbereich herzustellen. Man ist bestrebt, ein mikroelektronisches System von gegebenem Umfang auf möglichst kleiner Chipfläche unterzubringen; denn je größer das Chip, desto größer die Gefahr, dass es einen Kristalldefekt enthält, und desto geringer ist die Ausbeute bei seiner Fertigung. Die Chipfläche wird vor allem durch die Verdrahtung auf dem Chip bestimmt, da die Gatter im wesentlichen unter der Verdrahtung im Kristallinnern liegen. Es gilt daher beim Chipentwurf, die Gesamtlänge der Verdrahtung zu minimieren, und zwar aus zwei Gründen: Nicht nur die Chipfläche, sondern auch die Laufzeiten auf dem Chip werden dadurch verringert.

Die Weiterentwicklung der Mikroelektronik wurde stets von der Technologie der Halbleiterspeicher angeführt, da Speicher *regelmäßig strukturiert* sind: Die Speicherzellen sind matrixförmig angeordnet, Speicherchips weisen ein regelmäßiges Verdrahtungsmuster auf. Die Dichte der Bauelemente ist auf einem Speicherchip in der Regel höher als auf einem Mikroprozessor, da dessen Strukturen komplexer und damit unregelmäßiger sind. Regelmäßige Strukturen sind übersichtlicher beim Entwurf, mit höherer Ausbeute zu fertigen und von längerer Lebensdauer im Betrieb. Man sucht daher die Fortschritte der Speichertechnologie auch für Prozessor- und Logikschaltungen zu nutzen, indem man sie z.B. matrixförmig und damit möglichst regelmäßig strukturiert.

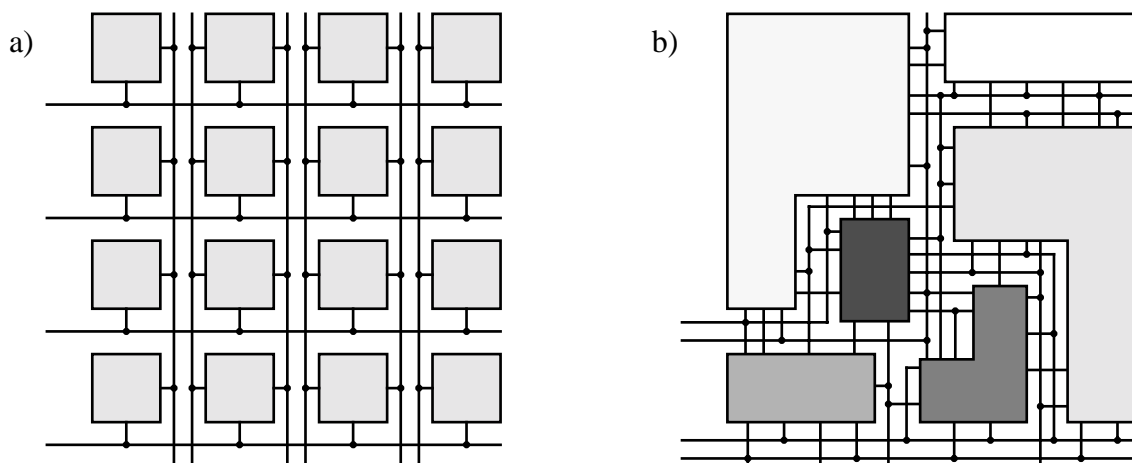


Bild 10.1: Grundstrukturen der Mikroelektronik; a) Speichermatrix; b) Mikroprozessor.

Einerseits setzt die Mikroelektronik hohe Investitionen an Gebäuden und Geräten voraus, so dass nur die automatisierte Herstellung großer Stückzahlen wirtschaftlich ist. Andererseits besteht ein großer Bedarf an anwendungsspezifischen Schaltungen, die unter Umständen nur in niedrigen Stückzahlen benötigt werden. Hier empfehlen sich *vordefinierte Strukturen*, die je nach Anwendung *personalisierbar* sind. Auch hier waren die Speicher die Vorreiter der Entwicklung: Alle Schaltfunktionen können durch Programmieren einer vorgefertigten Speichermatrix implementiert werden. Programmierbare Strukturen haben zudem den Vorteil der Flexibilität und Änderbarkeit, um Entwurfsfehler zu korrigieren und/oder neue Funktionen mit einzubeziehen.

10.1.2 Schaltungstechnologien

Elektronische Schaltungen und Systeme können mit zwei prinzipiell unterschiedlichen Schaltungstechnologien realisiert werden:

- Die bipolare Halbleitertechnik erlaubt die Herstellung von Transistoren mit NPN- oder PNP-Schichtstruktur sowie von PN-Dioden. Erstere sind aktive, letztere passive Bauelemente. Schaltungen mit bipolaren Bauelementen werden in der Regel statisch betrieben.
- Die unipolare Halbleitertechnik erlaubt die Herstellung von MOS-Transistoren mit N-Kanal oder P-Kanal. Beides sind aktive Bauelemente. Schaltungen mit MOS-Bauelementen können statisch oder, dank inhärenter MOS-Kapazitäten, auch dynamisch betrieben werden.

Aus der Fülle der Varianten, die bisher auf dem Gebiet der Schaltungstechnologien entwickelt wurden, sollen hier nur zwei einfache Vertreter vorgestellt werden. Denn es geht hier nicht um schaltungstechnische Einzelheiten, sondern um strukturell-topologische Zusammenhänge zwischen Logik- und Schaltungsebene. Auch die schaltungstechnisch hochinteressante CMOS-Technologie - es sei nur an die dynamische Domino-Logik erinnert - ginge in diesem Zusammenhang über die Zielsetzung des vorliegenden Abschnitts hinaus.

Bei der Realisierung logischer Gatter mit elektronischen Bauelementen muß zuerst die Zuordnung der binären Variablenwerte (0, 1) zu zwei elektrischen Pegeln definiert werden, die sich durch einen entsprechenden Störabstand deutlich voneinander unterscheiden. Trifft man für die Eingangs- und die Ausgangsvariablen dieselbe Zuordnung, so gibt es zwei Möglichkeiten:

- Positive Logik: Der höhere Signalpegel wird dem Wert 1, der niedrigere dem Wert 0 zugeordnet.
- Negative Logik: Der höhere Signalpegel wird dem Wert 0, der niedrigere dem Wert 1 zugeordnet.

Die folgenden Ausführungen setzen die *positive* Logik stillschweigend voraus. Als elementare Schaltungsbeispiele zeigt Bild 10.2 a) ein UND-Gatter mit bipolaren Dioden, b) ein ODER-Gatter in derselben Schaltungstechnik, c) ein NOR-Gatter mit NMOS-Transistoren, dazu die entsprechenden logischen Gattersymbole. Die Schaltalgebra lehrt, insbesondere ihr Hauptsatz, daß sich alle Schaltfunktionen und damit beliebige Schaltnetze allein mit diesen Grundgattern implementieren lassen.

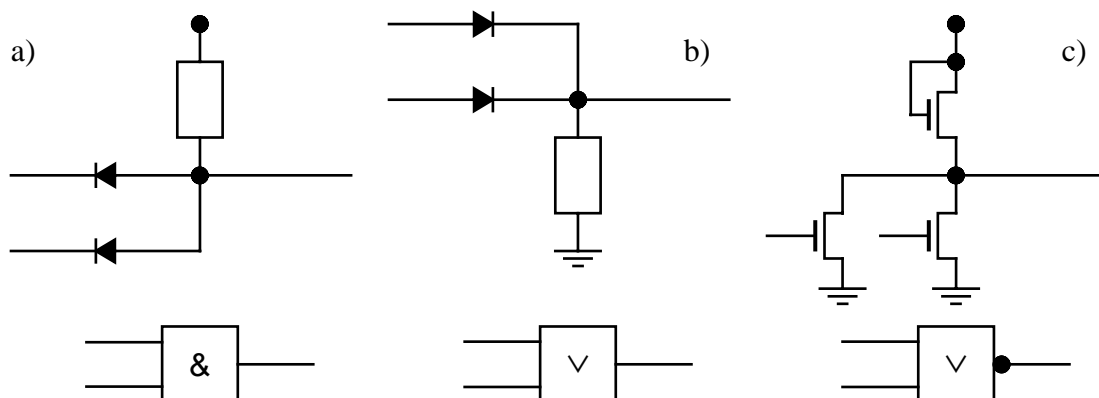


Bild 10.2: Schaltungstechnik; a) UND-Gatter mit Dioden, b) ODER-Gatter mit Dioden, c) NOR-Gatter mit NMOS-Transistoren.

Beispiel 10.1:

Gegeben sei folgender Strukturausdruck einer Schaltfunktion in disjunktiver Normalform:

$$\text{DNF} \quad y = (\bar{x}_2 \& x_1 \& \bar{x}_0) \vee (x_2 \& \bar{x}_1 \& x_0)$$

Das isomorphe zweistufige Schaltnetz zeigt das folgende Bild. Es besteht aus zwei UND-Gattern mit je drei Eingängen zur Implementierung der beiden Konjunktionen (&) und einem ODER-Gatter mit zwei Eingängen zur Implementierung der Disjunktion (\vee).

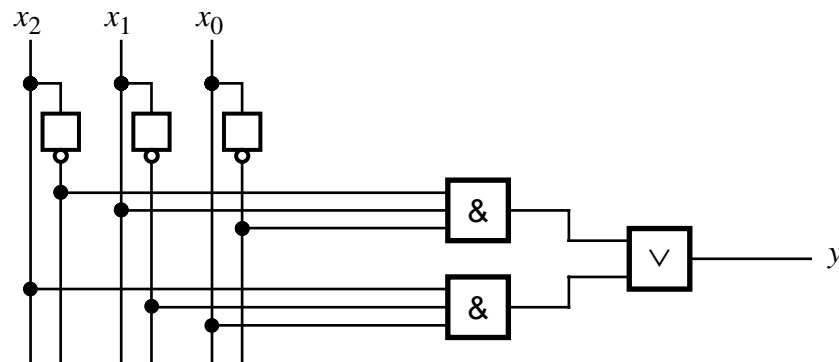


Bild 10.3: Zweistufiges Schaltnetz in disjunktiver Normalform (DNF)

Wendet man die DeMorgansche Regel auf die beiden Teilterme der obigen DNF an, so erhält man:

$$y = \overline{(\bar{x}_2 \& x_1 \& \bar{x}_0)} \vee \overline{(x_2 \& \bar{x}_1 \& x_0)} = \overline{(x_2 \vee \bar{x}_1 \vee x_0)} \vee \overline{(\bar{x}_2 \vee x_1 \vee \bar{x}_0)}$$

$$\bar{y} = (x_2 \vee \bar{x}_1 \vee x_0) \vee (\bar{x}_2 \vee x_1 \vee \bar{x}_0)$$

Damit lässt sich der gegebene Strukturausdruck auch durch ein zweistufiges Schaltnetz implementieren, das ausschließlich aus NOR-Gattern ($\bar{\vee}$) besteht. Der verneinte Funktionswert \bar{y} ist dabei durch einen einfachen Ausgangsinverter in den bejahten Funktionswert y umzuwandeln.

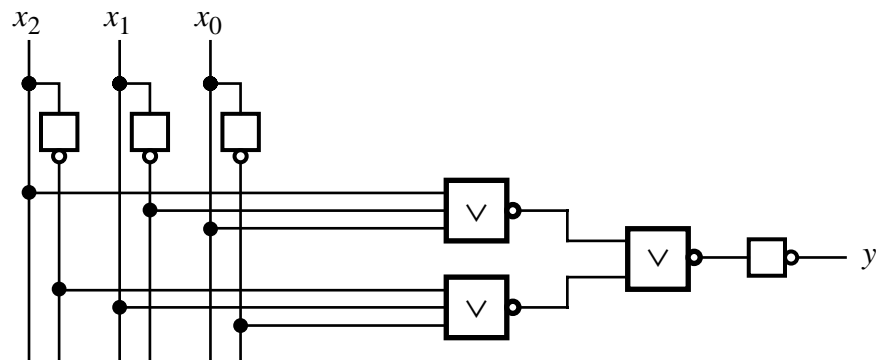


Bild 10.4: Zweistufiges NOR-Schaltnetz aus der DNF

a) Bipolare Schaltungstechnik

Die disjunktive Normalform (DNF) des obigen Beispiels (Bild 10.3) läßt sich in bipolarer Schaltungstechnik unter Verwendung der in Bild 10.2 a) und b) gezeigten Diodengatter direkt realisieren, wie Bild 10.5 zeigt. Die Schaltung besteht, soweit es die logischen Verknüpfungen betrifft, aus passiven Bauelementen, und sie wird statisch betrieben.

Interessant ist, daß sich die Diodenschaltung topologisch so umgestalten läßt, daß zur Realisierung der logischen Verknüpfungen zwei Matrizen entstehen (Bild 10.6):

- für die Konjunktionen (&) eine UND-Matrix, d.h. die Zeilen bilden logische Produkte P_i ,
- für die Disjunktion (\vee) eine ODER-Matrix, hier mit einer Spalte für den Funktionswert y .

Selbst dieses einfache Schaltbild zeigt, daß anwendungsspezifische Schaltfunktionen dadurch realisiert werden können, daß die Dioden in der UND-Matrix entsprechend der Spezifikation wahlweise an die bejahten oder die verneinten Eingangsvariablen x_i angeschlossen werden. Durch diese Maßnahme wird die vorstrukturierte Schaltung *personalisiert*.

b) MOS-Schaltungstechnik

Wird die disjunktive Normalform (DNF) desselben Beispiels wie oben gezeigt in eine zweistufige NOR-NOR-Form umgewandelt, so läßt sich die gegebene Schaltfunktion auch in MOS-Schaltungstechnik unter ausschließlicher Verwendung des in Bild 10.2 c) gezeigten NOR-Gatters realisieren. Die Schaltung in Bild 10.7 besteht aus aktiven, d.h. verstärkenden NMOS-Transistoren; sie wird hier statisch betrieben.

Auch die NMOS-Transistorschaltung läßt sich topologisch so umgestalten, daß zur Realisierung der logischen Verknüpfungen zwei Matrizen entstehen (Bild 10.8):

- für die Konjunktionen (&) eine UND-Matrix
- und für die Disjunktion (\vee) eine ODER-Matrix.

Die Namen der beiden Matrizen sind unabhängig davon, ob die logischen Verknüpfungen schaltungstechnisch durch eine echte UND-ODER-Schaltung oder, wie zumeist, durch eine zweistufige NOR-NOR-Schaltung realisiert werden.

Auch hier erfolgt die Personalisierung der Schaltung je nach anwendungsspezifischer Spezifikation dadurch, daß die Gatter der NMOS-Transistoren in der UND-Matrix wahlweise an die bejahten oder die verneinten Eingangsvariablen x_i angeschlossen werden.

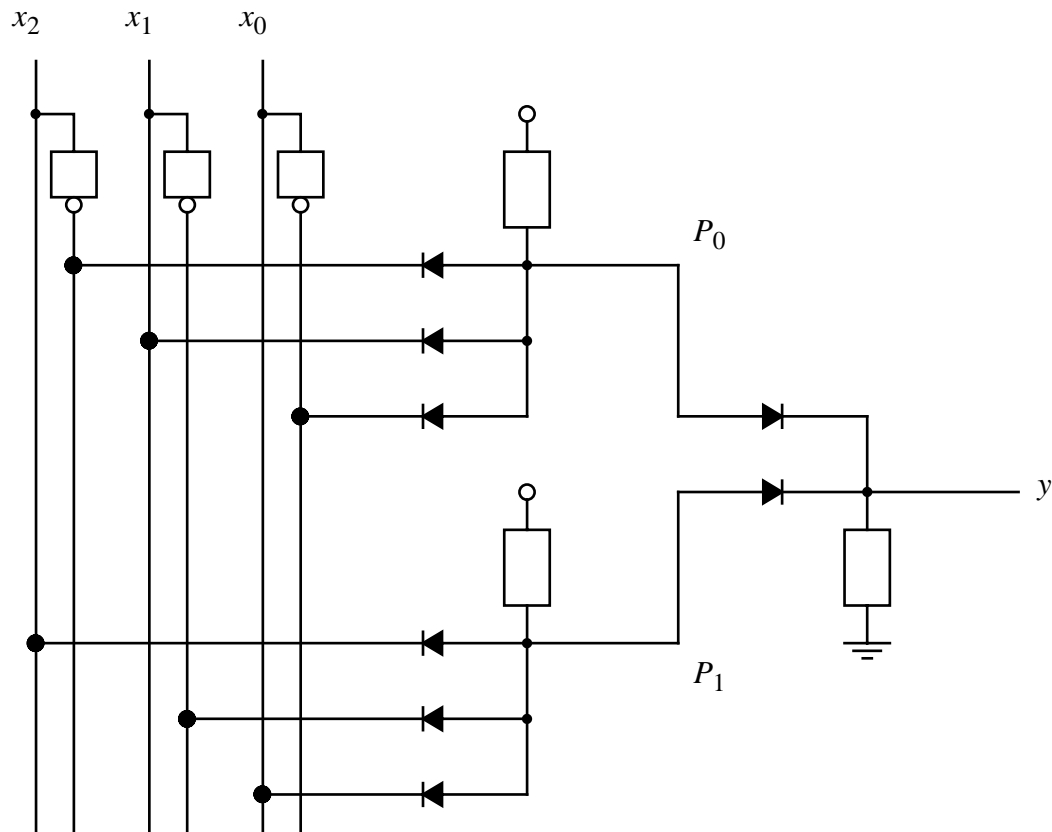
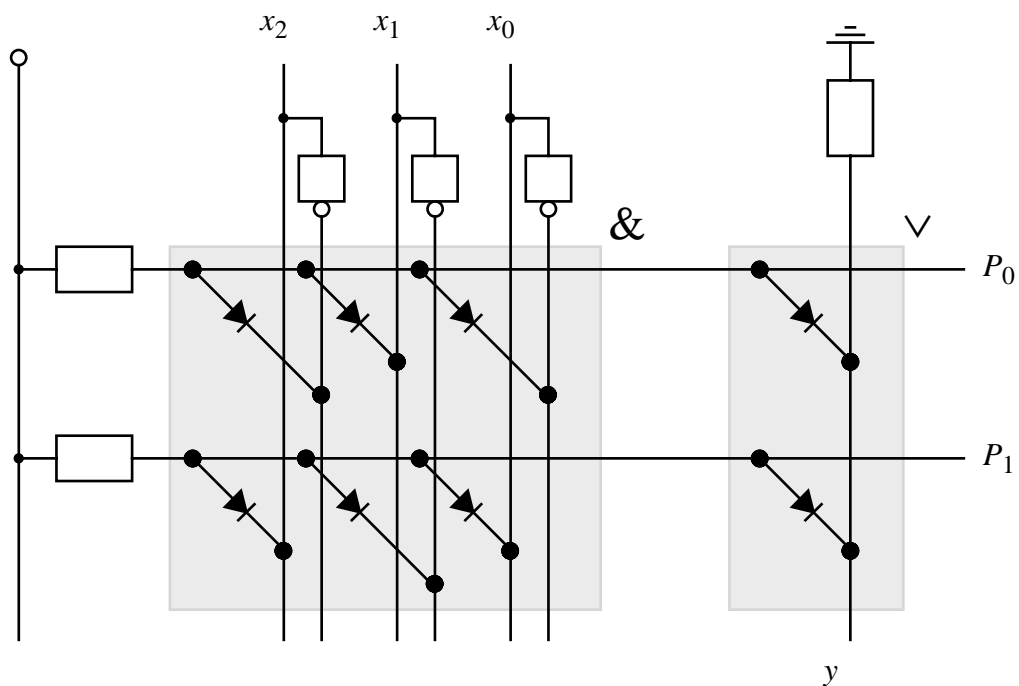


Bild 10.5: Realisierung des Schaltnetzes nach Bild 10.3 mit Diodengattern

Bild 10.6: Topologische Umgestaltung der Diodenschaltung nach Bild 10.5
& = UND-Matrix, ∨ = ODER-Matrix.

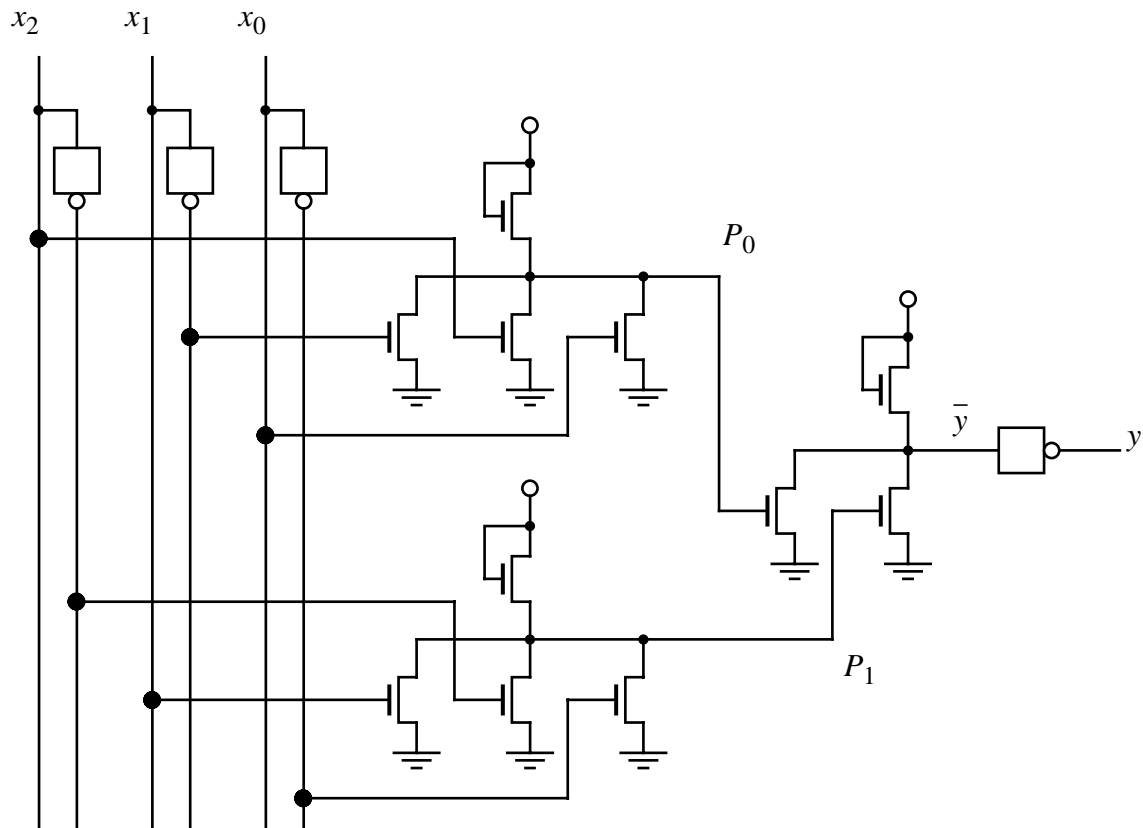


Bild 10.7: Realisierung des Schaltnetzes nach Bild 10.4 mit NMOS-Transistorgattern

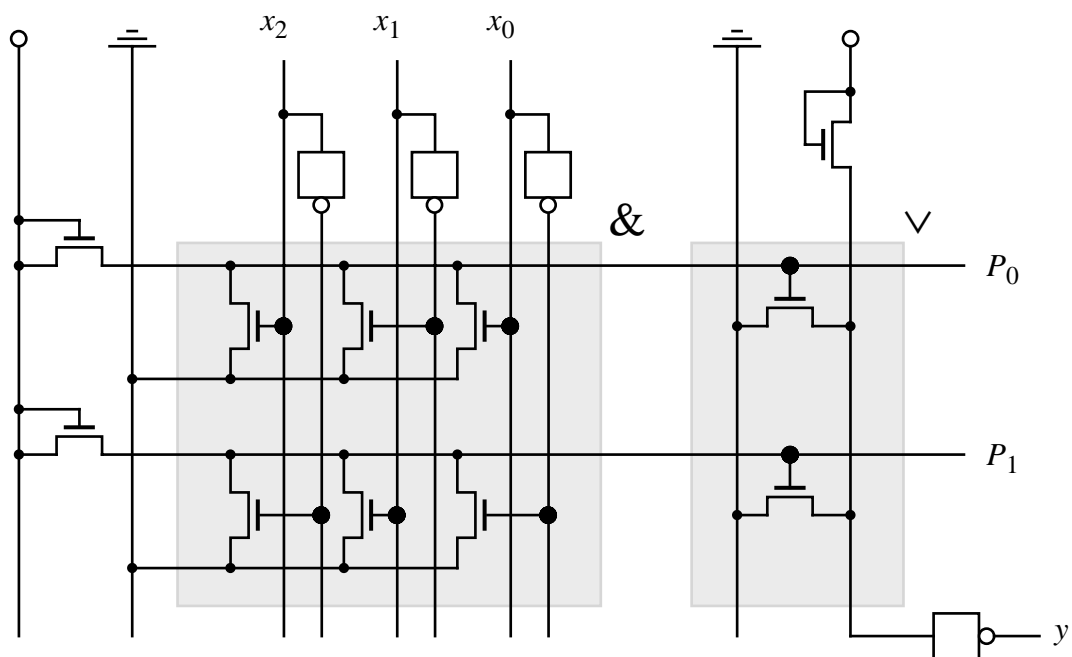


Bild 10.8: Topologische Umgestaltung der NMOS-Transistorschaltung nach Bild 10.7
& = UND-Matrix, v = ODER-Matrix.

10.1.3 Allgemeines Strukturschema

Matrixförmige Logikschaltungen nutzen das Potential der Mikroelektronik bei Entwurf, Herstellung und Betrieb besser aus als unregelmäßige Strukturen. Da sie *programmierbar* sind, gehören sie zur Klasse der anwendungsspezifischen integrierten Schaltungen („Application Specific Integrated Circuits“, ASIC). Indem man von schaltungstechnischen Einzelheiten absieht, können beide Schaltkreistechnologien, bipolar und MOS, durch das allgemeine Strukturschema im nachfolgenden Bild abstrahiert werden. Es besteht im wesentlichen aus einer UND-, gefolgt von einer ODER-Matrix, die beliebige logische Verknüpfungen in zweistufiger disjunktiver Form leisten.

Eine spezielle, anwendungsspezifische Schaltfunktion wird durch die *Personalisierung* der beiden Matrizen realisiert, wie es im Strukturschema durch die Anschlusspunkte angedeutet ist. Man kann die Personalisierung weiter abstrahieren und sie als *Programmierung* der Matrizen mit einem elementaren Binärcode betrachten, wie er nachfolgend angegeben wird. Man bezeichnet deshalb die in den folgenden Abschnitten zusammengestellten Varianten des allgemeinen Strukturschemas als Programmierbare Logische Bausteine („Programmable Logic Devices“, PLD).

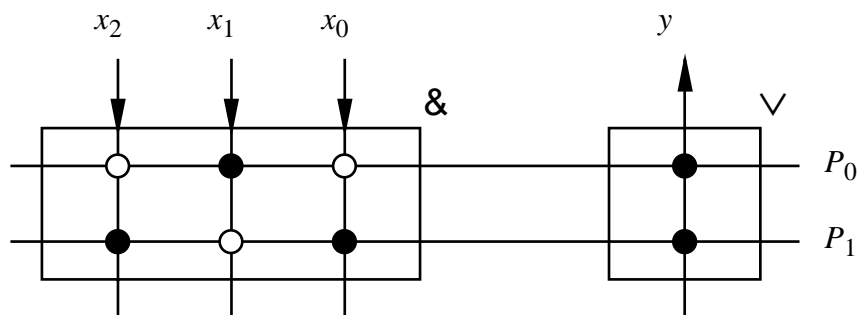


Bild 10.9: Allg. Strukturschema personalisierbarer Matrizen;
& = UND-Matrix; v = ODER-Matrix.

- Die UND-Matrix erzeugt die Konjunktion („logisches Produkt“) P_j der angeschlossenen Eingangsvariablen x_i . In dieser Matrix bedeutet:

		<u>Binärcode</u>
weißer Punkt:	die verneinte Eingangsvariable \bar{x}_i ist angeschlossen	0
schwarzer Punkt:	die bejahte Eingangsvariable x_i ist angeschlossen	1
kein Punkt:	die Eingangsvariable x_i ist nicht angeschlossen	-

- Die ODER-Matrix erzeugt die Disjunktion („logische Summe“) der angeschlossenen Konjunktionen P_j . In dieser Matrix bedeutet:

		<u>Binärcode</u>
schwarzer Punkt:	die Konjunktion P_j ist angeschlossen	1
kein Punkt:	die Konjunktion P_j ist nicht angeschlossen	0

Zur Realisierung von Bündelschaltfunktionen mit mehreren Funktionsspalten y_j kann die Anzahl der Spalten in der ODER-Matrix im obigen Bild entsprechend erhöht werden.

10.2 Programmierbare Schaltungen (PLDs)

10.2.1 Festwertspeicher (ROM)

Beispiel 10.2:

Ein binärer Volladdierer soll durch programmierbare Digitalbausteine, d.h. personalisierbare Schaltungen in Matrizenform realisiert werden. Er kann durch eine Wahrheitstabelle wie im nebenstehenden Bild spezifiziert werden. Dabei sind x_1 und x_0 die zusammen mit dem Übertrag c_0 aus der vorhergehenden Addition zu addierenden binären Variablen. Gebildet werden die Summe S und der neue Übertrag C_1 .

j	c_0	x_1	x_0	S	C_1
0	0	0	0	0	0
1	0	0	1	1	0
2	0	1	0	1	0
3	0	1	1	0	1
4	1	0	0	1	0
5	1	0	1	0	1
6	1	1	0	0	1
7	1	1	1	1	1

Bild 10.10: Funktionsbeschreibung eines binären Volladdierers

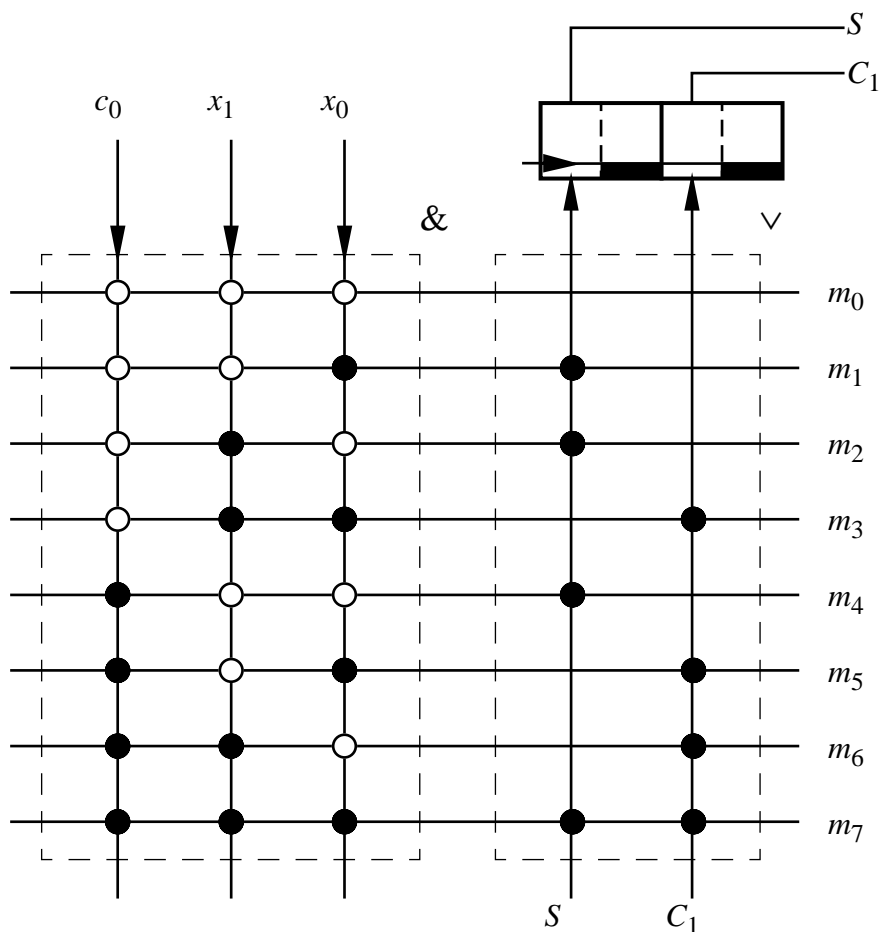


Bild 10.11: Festwertspeicher („Read Only Memory“, ROM);

& = vordefinierter 1-aus-8 Decoder, \vee = programmierbare Speichermatrix.

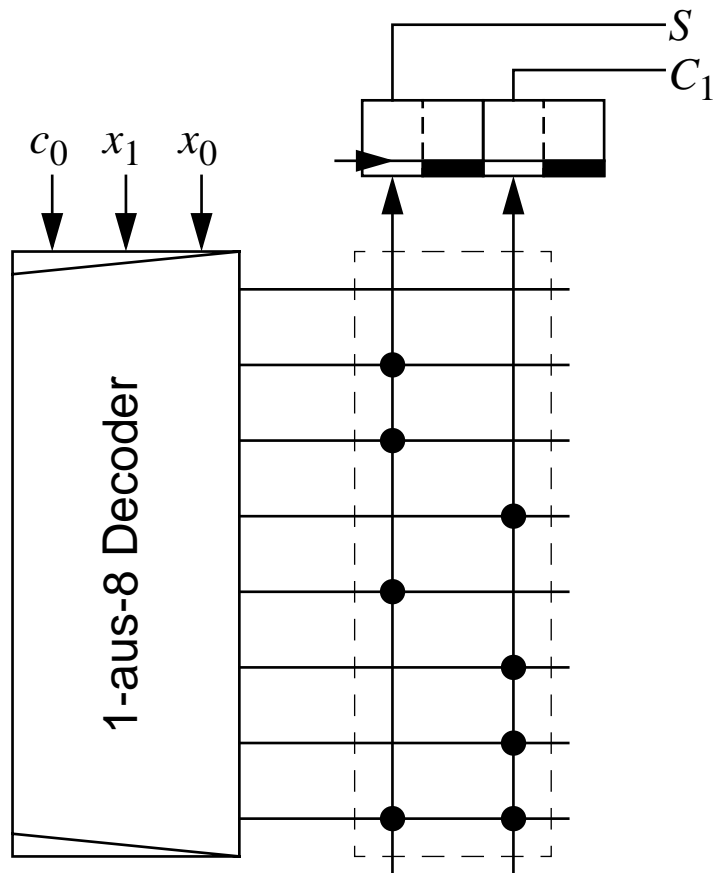


Bild 10.12: Ein-Ebenen-Adressierung eines Speichers mit wahlfreiem Zugriff (RAM)

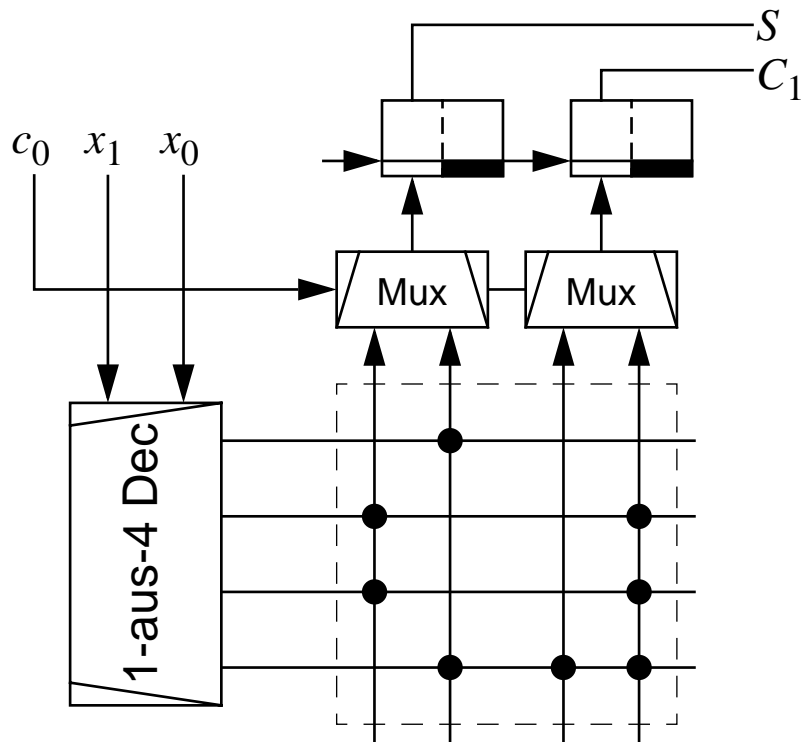


Bild 10.13: Zwei-Ebenen-Adressierung eines Speichers mit wahlfreiem Zugriff (RAM)

Personalisiert man die UND-Matrix wie im vorhergehenden Bild für $n = 3$ Eingangsvariable schematisch gezeigt, so werden durch ihre 2^n Zeilen sämtliche Minterme m_j gebildet, wobei $j = 0 \dots n-1$. Diese UND-Matrix entspricht dem 1-aus- n Decoder eines Speichers mit wahlfreiem Zugriff („Random Access Memory“, RAM) mit Einebenenadressierung. Wie das vorhergehende Bild ebenfalls zeigt, wird die ODER-Matrix isomorph zu den beiden Funktionsspalten der Wahrheitstabelle für die Summe S und den neuen Übertrag C_1 personalisiert („programmiert“); sie ist die eigentliche Speichermatrix.

Der 1-aus- n Decoder wird vom Hersteller vordefiniert, die Speichermatrix anwendungsspezifisch programmiert. Wird der Inhalt der Speichermatrix vom Hersteller nach Anwenderangaben fest personalisiert, so spricht man von einem *Festwertspeicher* („Read Only Memory“, ROM). Anmerkung: Die Bezeichnung RAM („Random Access Memory“) beschränkt sich in der Praxis auf vom Anwender im laufenden Betrieb *schreibbare* Speicher mit wahlfreiem Zugriff, obwohl auch ein sog. ROM, wie man seinem Strukturschema entnehmen kann, dank des 1-aus- n Decoders den wahlfreien Zugriff auf den Inhalt der Speichermatrix erlaubt.

Man beachte ferner, dass der Festwertspeicher in diesem Zusammenhang als rein kombinatorisches Schaltnetz wirkt. Ein sequentielles Schaltwerk entsteht daraus erst dann, wenn die beiden im obigen Strukturschema des ROM angedeuteten Flipflops am Ausgang der Speichermatrix auf den Eingang des Decoders rückgekoppelt werden.

10.2.2 Programmierbare UND-Matrixlogik (PAL)

Nun soll die gegebene Bündelschaltfunktion des Volladdierers minimiert werden, wobei das Funktionsbündel hier aus der Summe S und dem negierten Übertrag \bar{C}_1 besteht. (Die Negation erfolgt im Vorgriff auf die im nächsten Abschnitt vorgestellte Realisierung und kann durch Verwendung des negierenden Ausgangs eines Ausgangsflipflops wieder aufgehoben werden.) Die folgenden KV-Diagramme zeigen, dass sich nur die Schaltfunktion für \bar{c}_1 minimieren lässt.

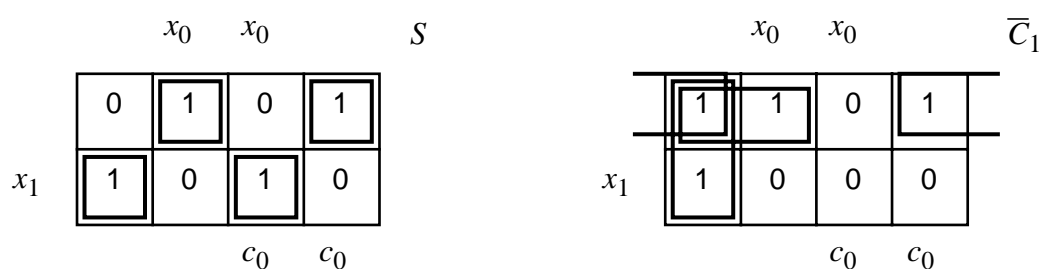


Bild 10.14: KV-Diagramme des Volladdierers mit Überdeckung durch Primblöcke

Man erhält durch Überdecken der Einsstellen mit Primblöcken in disjunktiver *Minimalform*:

$$\text{DMF } S = (\bar{c}_0 x_1 \bar{x}_0) \vee (c_0 x_1 x_0) \vee (\bar{c}_0 \bar{x}_1 x_0) \vee (c_0 \bar{x}_1 \bar{x}_0) = P_0 \vee P_1 \vee P_2 \vee P_3$$

$$\bar{C}_1 = (\bar{c}_0 \bar{x}_0) \vee (\bar{c}_0 \bar{x}_1) \vee (\bar{x}_1 \bar{x}_0) = P_4 \vee P_5 \vee P_6$$

Die zur DMF isomorphe Personalisierung zeigt das nächste Bild. In der UND-Matrix des Bausteins werden die gewünschten Konjunktionen („Produktterme“) P_j zeilenweise erzeugt. In einer durch den Hersteller des Bausteins vordefinierten ODER-Matrix ausreichender Größe werden sie dann disjunktiv verknüpft. Da nur die UND-Matrix anwendungsspezifisch personalisiert wird, während

ausgelegten Schaltung nicht nur die UND-, sondern auch die ODER-Matrix personalisiert werden kann. Diese Überlegungen führen zur *Programmierbaren Logischen Matrix* („Programmable Logic Array“, PLA).

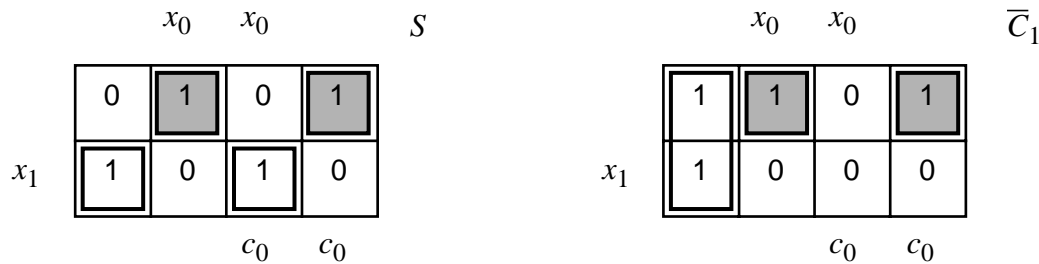


Bild 10.16: KV-Diagramme des Volladdierers mit Mehrfachausnutzung von Blöcken

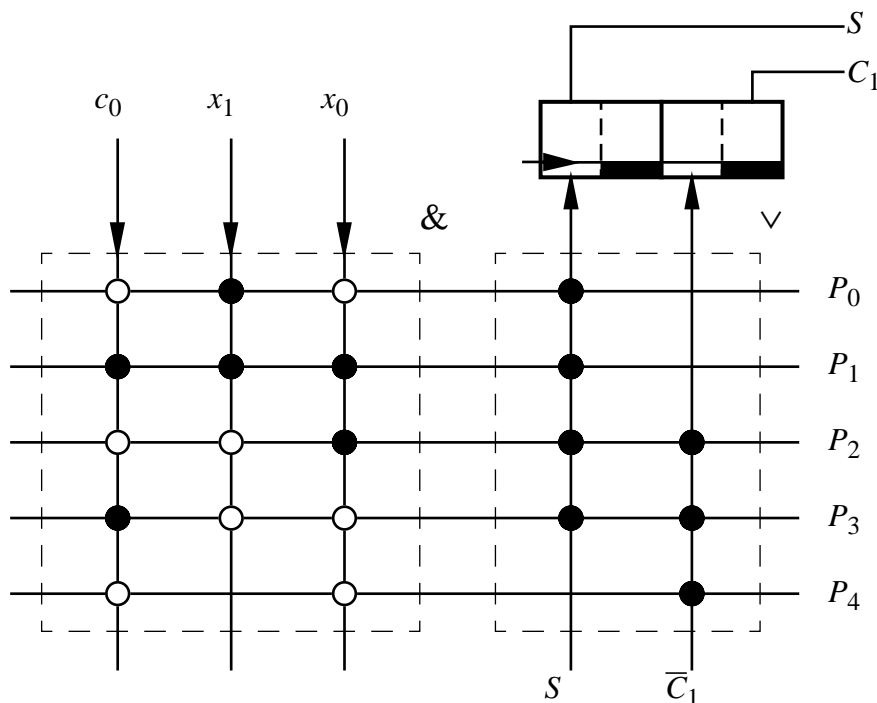


Bild 10.17: Programmierbare Logische Matrix („Programmable Logic Array“, PLA);
 $\&$ = programmierbare UND-Matrix, \vee = programmierbare ODER-Matrix.

Man erhält durch Überdecken der Einstellen mit mehrfach ausgenutzten Blöcken in disjunktiver *Normalform*:

$$\text{DNF} \quad S = (\bar{c}_0 \bar{x}_1 \bar{x}_0) \vee (c_0 x_1 x_0) \vee (\bar{c}_0 \bar{x}_1 x_0) \vee (c_0 \bar{x}_1 \bar{x}_0) = P_0 \vee P_1 \vee P_2 \vee P_3$$

$$\bar{C}_1 = (\bar{c}_0 \bar{x}_1 x_0) \vee (c_0 \bar{x}_1 \bar{x}_0) \vee (\bar{c}_0 \bar{x}_0) = P_2 \vee P_3 \vee P_4$$

Das Beispiel des Volladdierers zeigt das vorhergehende Bild. Die UND-Matrix des PLA erzeugt zeilenweise die gewünschten Produktterme P_j , in der ODER-Matrix wählt man die jeweils benötigten Produktterme aus, d.h. beide Matrizen sind anwendungsspezifisch zu personalisieren. Im Vergleich zum PAL ergibt sich beim PLA eine Einsparung von zwei weiteren Produkttermen durch deren Mehrfachausnutzung.

10.3 Integrierte Schaltungen (ICs)

Alle digitalen Systeme werden „anwendungsspezifisch“ entwickelt, d.h. sie sollen eine bestimmte Systemfunktion erfüllen. Sie werden nach dem Stand der Technik mit auf Siliziumchips integrierten Schaltungen aufgebaut („Integrated Circuits“, ICs).

Mikroprozessoren & Speicherchips

Die anwendungsspezifische Systemfunktion wird in *Software* implementiert: Sie beruht im Bitmuster eines Speichers, das vom Mikroprozessor interpretiert wird. Mikroprozessor- und Speicher-ICs können universell eingesetzt werden.

Integrierte Logikschaltungen

Die anwendungsspezifische Systemfunktion wird in *Hardware* realisiert: Sie beruht im Verdrahtungsmuster einer oder mehrerer integrierter Schaltungen. Logik-ICs werden speziell entwickelt und eingesetzt.

Dieselbe Systemfunktion kann also entweder mit Mikroprozessoren und Speichern („Softwarelösung“) oder mit Logikschaltungen („Hardwarelösung“) ausgeführt werden. Der Unterschied besteht dabei nicht nur in den Schaltungsstrukturen, sondern vor allem in der angewandten *Entwurfsmethodik*, um die anwendungsspezifische Funktion zu verwirklichen.

Derzeit wird eine gemeinsame Entwurfsmethodik für komplexe digitale Systeme durch „Hardware/Software-Codesign“ erforscht und entwickelt.

10.3.1 Integrierte Schaltungsstrukturen

Betrachtungen zur Wirtschaftlichkeit

- Die Mikroelektronik setzt hohe Investitionen an speziellen Gebäuden und hochkomplizierten Geräten voraus, so daß eine Chip-Herstellung nur in großen Stückzahlen wirtschaftlich ist.

Speicher und Mikroprozessoren sind *universell verwendbar* und daher in großen Stückzahlen herstellbar.

- Es besteht jedoch auch ein großer Bedarf an unterschiedlichen, anwendungsspezifischen Logikschaltungen, die jeweils oft nur in geringer Stückzahl benötigt werden.

Hier empfehlen sich *vordefinierte Strukturen*, die in hoher Stückzahl hergestellt werden können und die je nach Anwendung personalisierbar sind.

Integrierte Logikschaltungen

Logik-ICs lassen sich nach ihrem Integrationsgrad, ihrer Entwurfsmethodik und ihrer Flexibilität für unterschiedliche Anwendungen in drei Klassen einteilen:

- Voll-kundenspezifische Logik-ICs,
- Semi-kundenspezifische Logik-ICs,
- Standard Logik-ICs,

wobei grundsätzlich gilt: Je umfassender und komplexer ein Logik-IC aufgebaut, d.h. je komplizierter seine Gesamtfunktion ist, desto seltener kann er in unterschiedlichen Anwendungen eingesetzt werden - und umgekehrt.

10.3.2 Standard Logik-ICs

Integrationsgrad und Einsetzbarkeit

Es handelt sich um Schaltungen niedrigen Integrationsgrades ($1 \cdot 10^2$ Gatter pro Chip) in standardisierter Ausführung, um die hohen Stückzahlen zu erreichen. Sie sind in unterschiedlichen Anwendungen flexibel einsetzbar.

Auf dem Chip sind universell verwendbare NAND-Gatter und Flipflops, aber ggf. auch einzelne Register, Decoder, Multiplexer, Addierer und dergl. vorhanden. In einem Modul sind 1..3 Standardchips untergebracht, d.h. das Modulgehäuse („IC“) enthält die funktionellen Elemente für den Schaltungsentwurf auf Logik- und/oder RT-Ebene.

Entwurfsmethodik und Kosten

Der Anwender muß hier nur die System- und die Logikentwicklung selbst durchführen und er kann sich auf das Platinen-Layout beschränken. Es entfällt für ihn nicht nur die Schaltungsentwicklung, sondern auch das Chip- und das Modul-Layout.

Entwicklungszeit und -kosten können relativ niedrig gehalten werden. Wegen des niedrigen Integrationsgrades von Standard-ICs ist die erreichbare Systemkomplexität eingeschränkt, da das Platzangebot und der maximale Energiebedarf vorgegeben sind, und die Fertigungskosten eines Systems sind relativ hoch.

Technologie

Standard-ICs werden bipolar in „Transistor-Transistor Logic“ (TTL) oder in „Complementary MOS“ (CMOS) hergestellt.

Geometrische Konstruktion

- Partitionierung der Platine in Modulplätze
- Platzierung von Standard-Modulen („Bestückung“)
- Verdrahtung zwischen den Modulen („Entflechtung“)

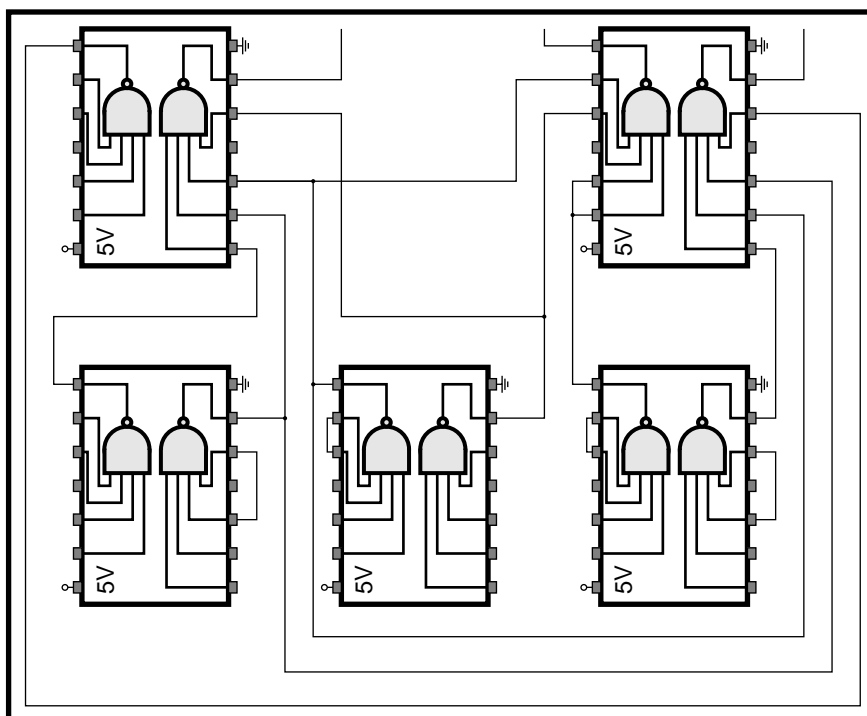


Bild 10.18: Schaltung mit Standard Logik-ICs

10.3.3 Voll-kundenspezifische Logik-ICs

Integrationsgrad und Einsetzbarkeit

Es handelt sich um maßgeschneiderte Schaltungen extrem hohen Integrationsgrades ($10^4 \dots < 10^6$ Gatter pro Chip), die speziell für eine einzige Anwendung entwickelt werden, falls sie dafür in großer Stückzahl benötigt werden ("full-custom IC").

Entwurfsmethodik und Kosten

Der Anwender muß hier alle Phasen des Entwurfs selbst durchführen: System-, Logik- und Schaltkreisentwicklung, ferner das Chip-, Modul- und Platinen-Layout sowie ggf. die geometrische Konstruktion darüber hinausgehender, globaler Packungseinheiten.

Der Entwurf voll-kundenspezifischer ICs ist nur mit umfassender Rechnerunterstützung möglich. Gleichwohl sind die Entwicklungszeiten relativ lang und die Entwicklungskosten relativ hoch gegenüber den relativ niedrigen Fertigungskosten des auf wenigen, meist nur einem einzigen Chip integrierten Systems, so daß sich der hohe Entwicklungsaufwand nur bei sehr großen Stückzahlen ($10^5 \dots 10^6$) lohnt.

Technologie

Voll-kundenspezifische ICs werden vorzugsweise in "Complementary MOS" (CMOS) oder bipolar in "Emitter Coupled Logic" (ECL) hergestellt.

Geometrische Konstruktion

- Partitionierung des Chips („Floorplanning“)
- Platzierung selbst definierter Funktionszellen
- Verdrahtung zwischen den platzierten Zellen

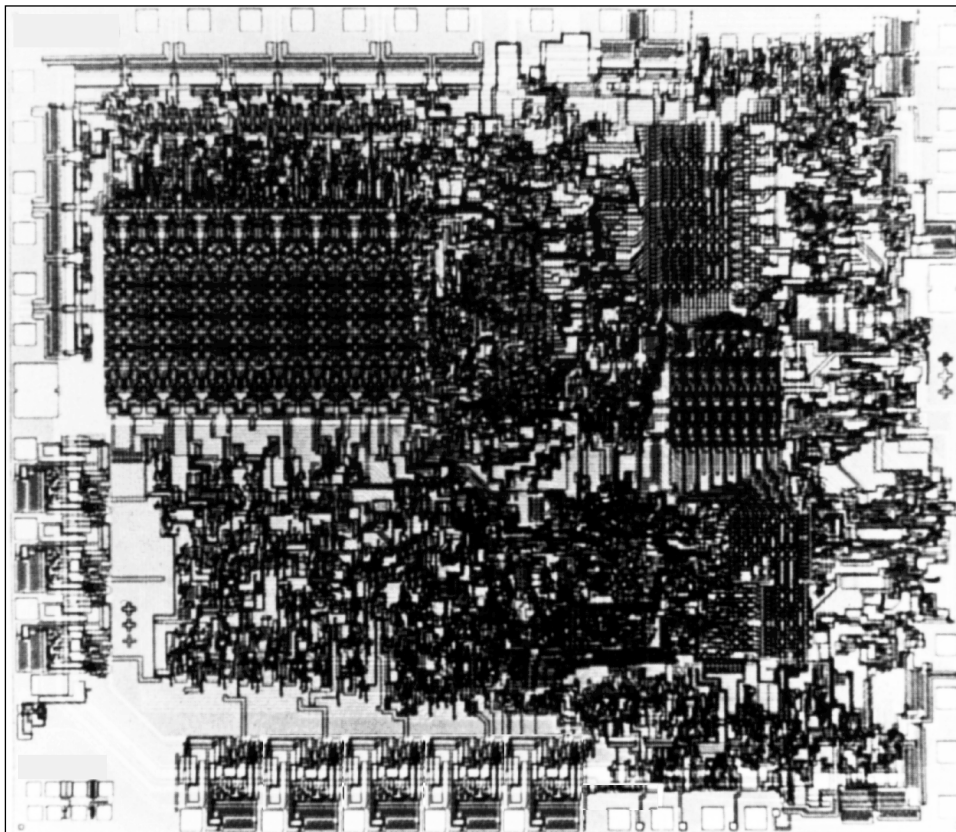


Bild 10.19: Voll-kundenspezifischer Logik-IC

10.3.4 Semi-kundenspezifische Logik-ICs

Integrationsgrad und Einsetzbarkeit

Es handelt sich um teilweise vordefinierte Schaltungen hohen Integrationsgrades ($10^2 \dots 10^5$ Gatter pro Chip), die in unterschiedlichen Anwendungen flexibel eingesetzt und dank vorgegebener Layoutstrukturen trotzdem in großer Stückzahl gefertigt werden können („semi-custom IC“). Sie werden auch als anwendungsspezifische integrierte Schaltungen bezeichnet („Application Specific Integrated Circuits“, ASIC).

Entwurfsmethodik und Kosten

Der Anwender muß hier nur die System- und Logikentwicklung selbst durchführen, ferner das Chip-Layout durch Platzierung und Verdrahtung von Funktionszellen erstellen, ggf. auch das Modul- und/oder das Platinen-Layout. Dank der vorgegebenen Layoutstruktur und der vom Hersteller vordefinierten Funktionszellen entfällt für ihn die Schaltungsentwicklung und das Layout auf Schaltungsebene.

Bei geeigneter Rechnerunterstützung werden die Entwicklungszeiten gegenüber voll-kundenspezifischen ICs verkürzt und die Entwicklungskosten deutlich gesenkt. Mit semi-kundenspezifischen ICs können Systeme hoher Komplexität mit wenigen Chips realisiert werden, wobei dank höherer Ausbeute auch die Fertigungskosten niedriger sind.

Technologie

Semi-kundenspezifische ICs werden in „Complementary MOS“ (CMOS), bipolar in „Emitter Coupled Logic“ (ECL) oder „Transistor-Transistor Logic“ (TTL) hergestellt.

Geometrische Konstruktion

- Partitionierung des Chips („Floorplanning“)
- Platzierung vordefinierter Funktionszellen
- Verdrahtung zwischen den platzierten Zellen

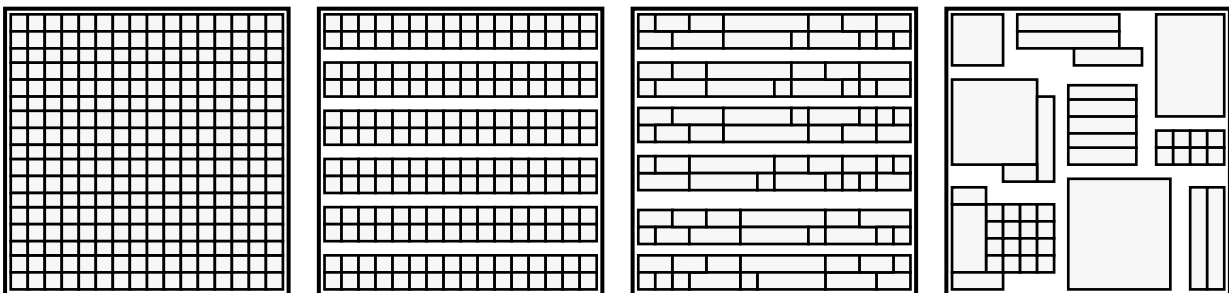


Bild 10.20: Semi-kundenspezifische Logik-ICs (Beispiele)

Wegen ihrer vielseitigen Verwendbarkeit werden nachfolgend vor allem die semi-kundenspezifischen Logik-ICs behandelt. Sie lassen sich weiter unterteilen:

- Gate-Arrays mit Matrix-Struktur
- Standardzellen-ICs mit Zeilen-Struktur
- Makrozellen-ICs mit Manhattan-Struktur

10.3.5 Gate-Arrays

Technologie und Zellbibliothek

Vom Technologieanbieter werden universell verwendbare *Funktionszellen*, die auf *Basiszellen* des gewählten Gate-Arrays abbildbar sind, in einer technologiespezifischen „Zellbibliothek“ für den Anwender bereitgestellt.

Dank des vordefinierten Master-Layouts können (fast) alle Fertigungsschritte eines Gate-Arrays vorab durchgeführt werden. Nach Abschluß des Entwurfs sind meist nur noch die entsprechenden Verdrahtungsmuster herzustellen. Die Breite der Verdrahtungskanäle ist konstant.

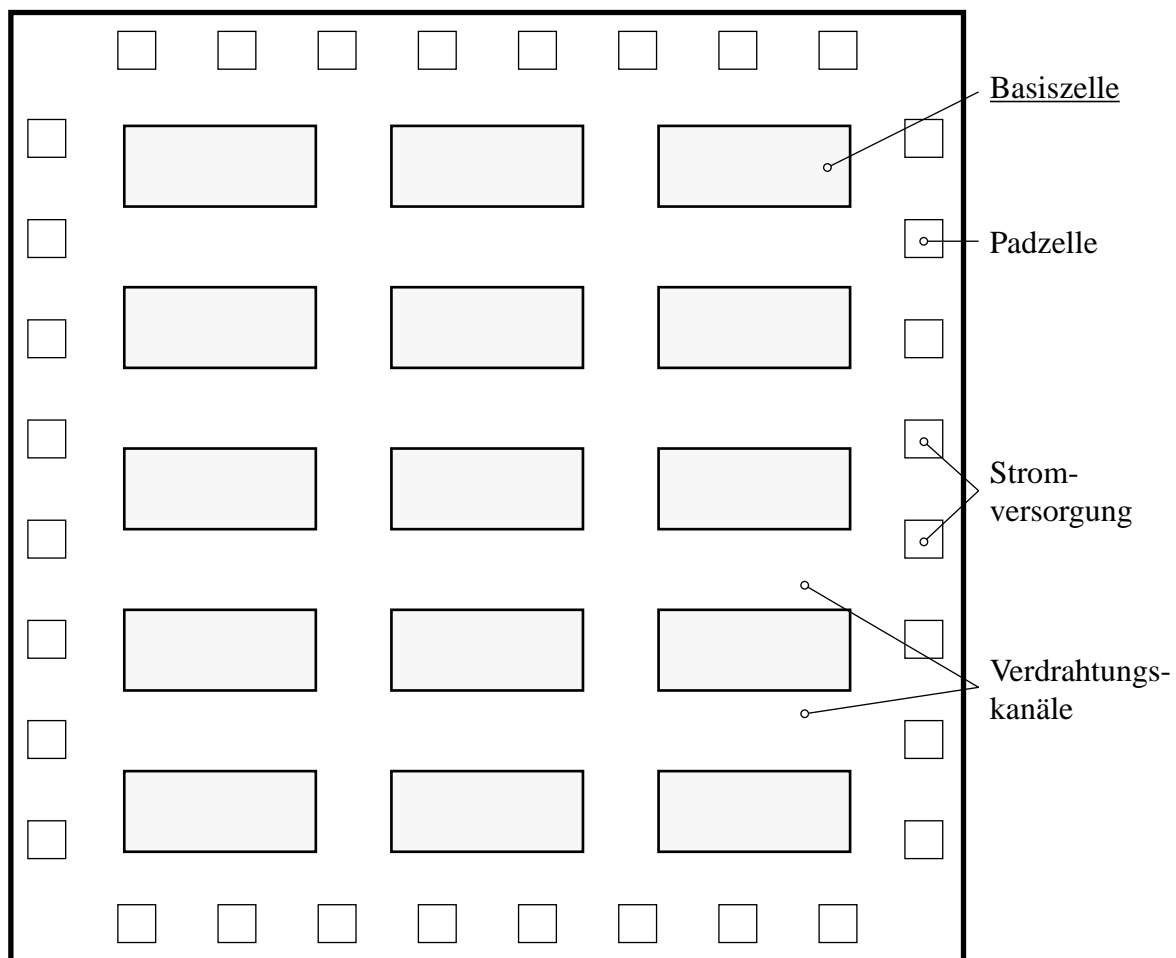


Bild 10.21: Master-Layout mit Matrix-Struktur

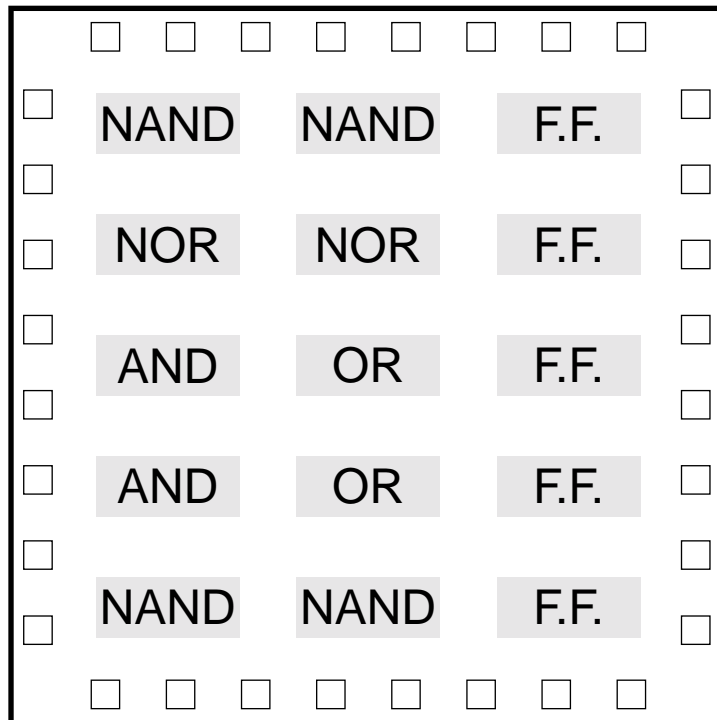


Bild 10.22: Platzierung durch „Intrazell-Verdrahtung“ auf einem Gate-Array mit Matrixstruktur

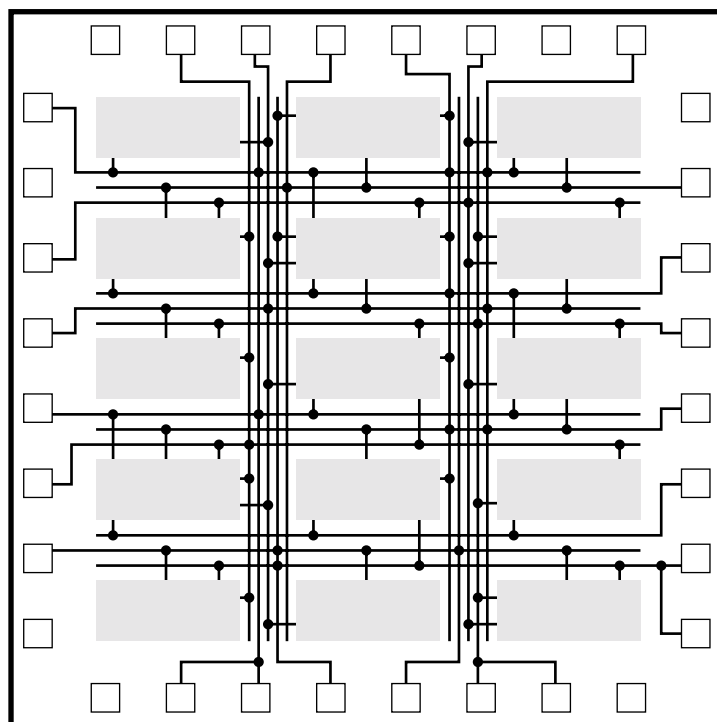


Bild 10.23: Verdrahtung durch „Interzell-Verdrahtung“ auf einem Gate-Array mit Matrixstruktur

10.3.6 Standardzellen-ICs

Technologie und Zellbibliothek

Ein Satz universell verwendbarer *Standardzellen* wird vom Technologieanbieter entworfen und dem Anwender als technologiespezifische „Zellbibliothek“ zur Verfügung gestellt.

Nach Abschluß des Entwurfs eines Standardzellen-IC müssen *alle* Fertigungsschritte durchgeführt werden. Die Breite der Verdrahtungskanäle ist variabel.

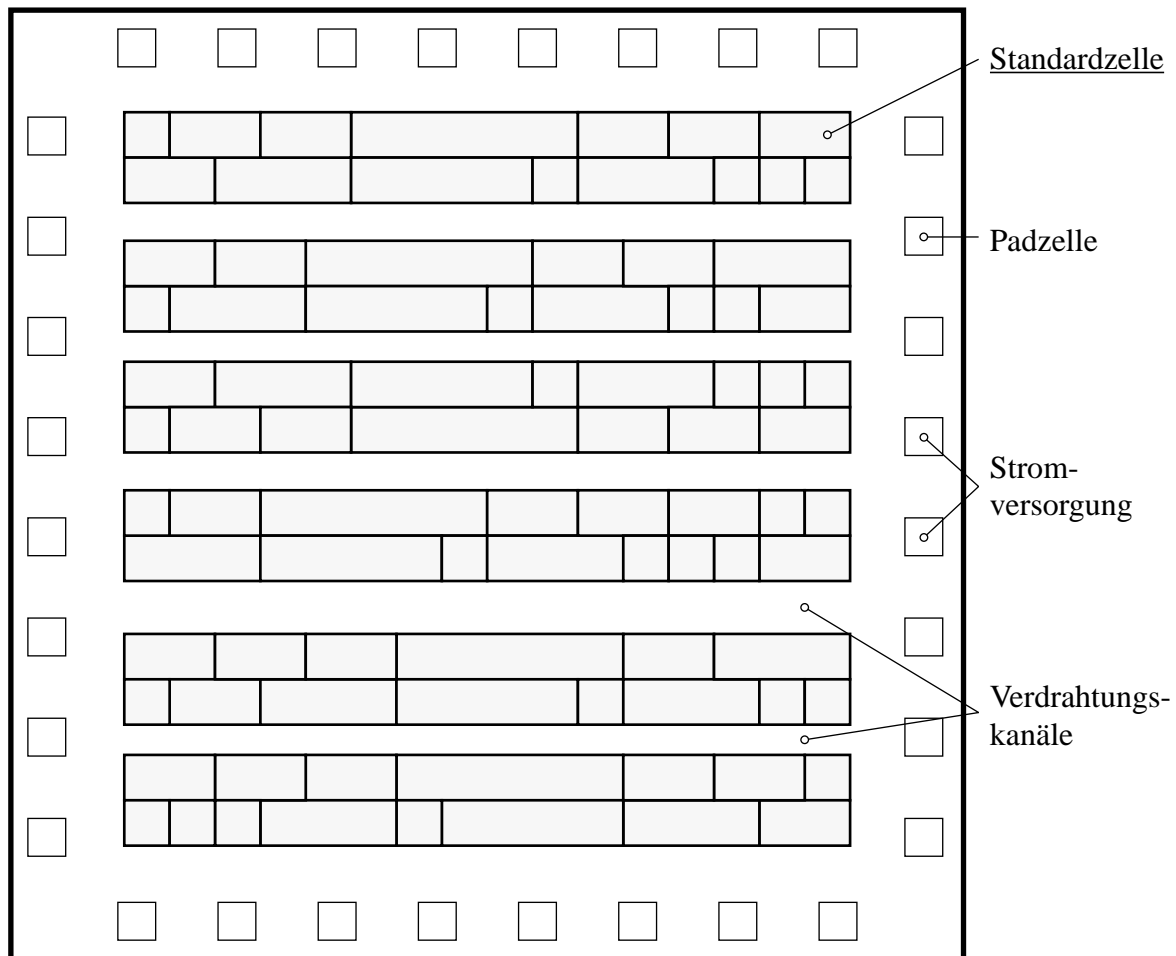


Bild 10.24: Layout mit Zeilen-Struktur

10.3.7 Makrozellen-ICs

Technologie und Zellbibliothek

Für jeden Anwendungsfall werden die meisten *Makrozellen* individuell entworfen, gegebenenfalls aus universell verwendbaren, einfacheren Zellen.

Nach Abschluß des Entwurfs eines Makrozellen-IC müssen *alle* Fertigungsschritte durchgeführt werden. Die Breite der Verdrahtungskanäle ist unterschiedlich.

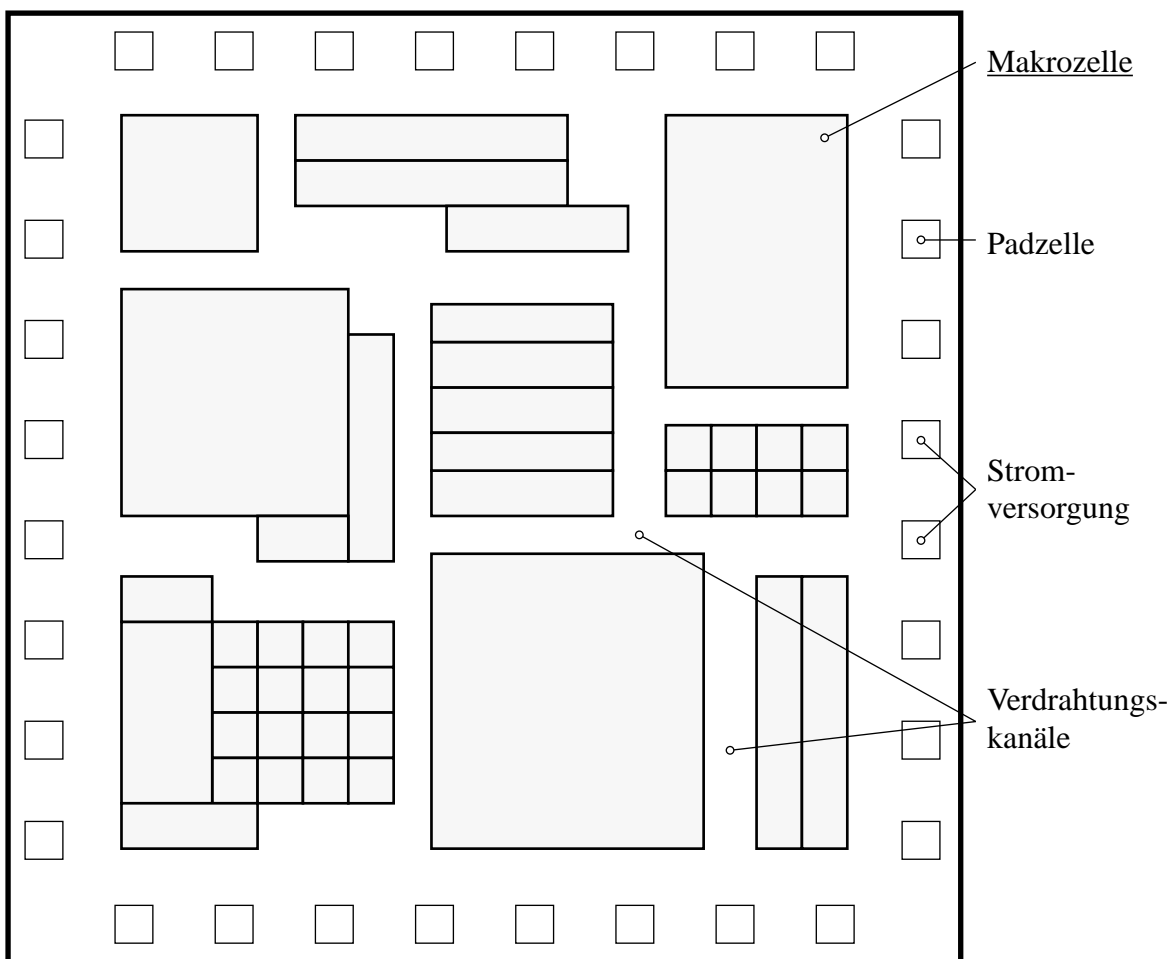


Bild 10.25: Layout mit Manhattan-Struktur