Elektronik für Physiker

Scriptum zur Vorlesung im Wintersemester 1998/99

U. Straumann

Physikalisches Institut Universität Heidelberg

Einleitung

Die Elektronik spielt im Leben jedes erfolgreichen Experimentalphysikers eine zentrale Rolle. Praktisch alle physikalischen Messungen werden vom Experimentator - anstelle direkter Beobachtung mit einem seiner fünf Sinne - mit Hilfe von Sensoren durchgeführt, die die physikalische Messgrösse in eine elektrischen Spannung oder Ladung transformiert. *(Sechste und höhere Sinne ("Intuition") sind allerdings nach wie vor unverzichtbare Hilfsmittel für ein erfolgreiches Experimentieren, besonders auch bei Verwendung von elektronischen Geräten.)* Diese elektrischen Messignale müssen verstärkt, gefiltert, digitalisiert, komprimiert und einem Rechner zugeführt werden, wo die Daten schliesslich protokolliert und analysiert werden. In dieser Vorlesung sollen alle diese Schritte der Messkette behandelt werden.

"Messen" bedeutet bekanntlich "Fehler machen". Ein zentrales Anliegen der Elektronikausbildung eines Physikers muss es deshalb auch sein, das Verständnis für die in elektronischen Messschaltungen auftretenden Messfehler zu wecken und zu fördern.

Experimentieren heisst aber auch, Naturvorgänge unter wohldefinierten Umweltbedingungen zu beobachten. Neben der Messung stellen deshalb auch die Steuerung und Regulierung dieser Umweltbedingungen ein weites Anwendungsgebiet für elektronische Schaltungen dar, die vom Physiker verstanden, richtig eingesetzt und gelegentlich auch selbst entwickelt werden müssen.

Experimentalphysiker verwenden fertige elektronische Schaltungen, oder sie legen die Spezifikationen für die Entwicklung von neuen Geräten fest. Oft stehen sie auch vor dem Problem, dass mehrere Geräte mit den richtigen Funktionen zwar zur Verfügung stehen, aber signalmässig nicht richtig zu einander passen. Dann muss ein Interface gebastelt werden.

Das Ziel dieser einsemestrigen Einführung ist es dementsprechend, die Hörer zu befähigen,

1. einschlägige elektronische Schaltungen funktionsmässig und numerisch zu diskutieren und ihre Genauigkeiten und Grenzen abzuschätzen.
2. für Neuentwicklungen realistische Spezifikationen zu erarbeiten.
3. kompetente Gesprächspartner von Elektronikingenieuren zu sein.

Diese Vorlesung richtet sich an Studenten in der mittleren Ausbildung und soll die Grundlage für die experimentelle Tätigkeit legen, die in der Regel mit der Diplomarbeit beginnt. Sie ist grundsätzlich praktisch orientiert. So werden viele technische Einzelheiten behandelt, oder es wird auf entsprechende Literatur und Tabellen verwiesen.

Die Vorlesung legt etwa gleichviel Gewicht auf digitale und analoge Elektronik. Der Aufbau und die Funktionsweise von Computern wird jedoch nicht direkt behandelt (siehe seperate Veranstaltungen). Auf der anderen Seite wird ebenfalls die physikalische Funktionsweise von Halbleiterbauelementen nur sehr vereinfachend diskutiert, wie sie für den Anwender relevant sind.

Simulationsprogramme *(SPICE)* und rechnergestützte Entwicklungshilfsmittel (Simulation, Synthese, Layout usw.) werden entsprechend ihrer Bedeutung vorgestellt.

Inhaltsverzeichnis

Kapitel 1

Strom, Spannung, Widerstand

Das Ohm'sche Gesetz stellt die zentrale Grundlage aller elektronischen Anwendungen dar. In seiner elementaren Form $U = R \cdot I$ gilt es jedoch nur für zeitlich konstante Situationen immer. In diesem Kapitel werden als Repetition die *komplexe* Darstellung von zeitabhängigen Strömen und Spannungen, sowie Widerstände dargestellt als komplexe Zahlen ausführlich behandelt. Deren praktische Realisierung besteht aus passiven Bauelementen in Form von Widerständen, Kondenstoren und Induktivitäten.

Als weitere wichtige Elemente in der Elektronik werden darauf Strom- und Spannungsquellen diskutiert, jeweils in ihrer idealisierten und realen Form.

Mit diesen Begriffen können schliesslich die Kirchhoff'schen Regeln aufgestellt werden, welche für die konkrete Berechnung von Schaltungen eine zentrale Rolle spielen.

1.1 Komplexe Widerstände, passive Bauelemente

1.1.1 Komplexe Spannungen und Ströme

Spannungen und Ströme sind im allgemeinen beliebige zeitabhängige Funktionen $U(t)$, $I(t)$. Diese können nach dem Satz von Fourier als (evt. unendlich dicht liegende) Summen von harmonischen Funktionen dargestellt werden. Diese werden als komplexe Funktionen geschrieben ($i := +\sqrt{-1}, \quad i^2 = -1, \quad i^3 = -i, \quad i^4 = 1$):

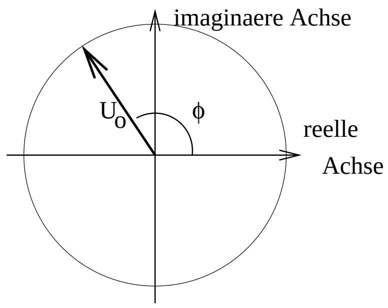
$$\widetilde{U(t)} = U_0 \cdot e^{i(\omega t + \varphi_u)} = U_0 e^{i\varphi_u} e^{i\omega t} \tag{1.1}$$

$$\widetilde{I(t)} = I_0 \cdot e^{i(\omega t + \varphi_I)} = I_0 e^{i\varphi_I} e^{i\omega t} \tag{1.2}$$

Dabei ist der Realteil der komplexen Funktion der wirkliche physikalische Strom (bzw. Spannung). $\omega = 2\pi\nu$ heisst die Kreisfrequenz und φ_u (φ_I) die Phase der Spannung (des Stromes). Mit Hilfe der Eulergleichung

$$\begin{aligned} e^{ix} &= \cos x + i \cdot \sin x & (1.3)\\ \Rightarrow \text{Realteil}(e^{ix}) &= \cos x \\ \text{Imaginärteil}(e^{ix}) &= \sin x \end{aligned}$$

kann der physikalische Realteil der Funktion berechnet werden.



Während die Beträge U_0 und I_0 der komplexen Funktionen den reellen Amplituden (*Scheitelwerte*) entsprechen, stellen die Grössen $U_0 e^{i\varphi_u}$ bzw. $I_0 e^{i\varphi_I}$ die Spannung und Stromstärke zur Zeit null dar und werden deshalb *komplexe Amplituden* genannt. Diese werden wie in der nebenstehenden Zeichnung oft in der komplexen Ebene als *Zeiger* dargestellt, wobei der zeitabhängige Teil $e^{i\omega t}$ weggelassen wird. Man spricht in diesem Fall auch von *Operatoren*.

Der *Effektivwert* von Strom oder Spannung wird durch

$$U_{eff} = \sqrt{\frac{1}{T}\int_0^T U^2(t)dt} \tag{1.4}$$

definiert. Es handelt sich also um den quadratischen Mittelwert, der auch mit r.m.s. (*root mean square*) oder mit σ bezeichnet wird. Der Effektivwert ist relevant für die z.B. in elektronischen Bauteilen entstehende Wärme durch die Verlustleistung, die durch den zeitlichen Mittelwert des Produktes aus Strom und Spannung $P = \overline{U \cdot I} = \overline{U^2/R}$ bestimmt ist. Für sinusförmige Wechselströme wird

$$P = U_{eff} \cdot I_{eff} \cdot \cos\varphi \tag{1.5}$$

wobei φ die Phasendifferenz zwischen Strom und Spannung darstellt.

Das Verhältnis zwischen Scheitelwert und Effektivwert wird als *Scheitelfaktor* k_s (englisch *crest factor*) bezeichnet:

$$k_s := \frac{U_0}{U_{eff}} \tag{1.6}$$

Für sinusförmige Wechselspannungen ist $k_s = \sqrt{2}$. In Wechselspannungs–Leistungsversorgungen, wie zum Beispiel dem öffentlichen Netz, werden jeweils die Effektivwerte angegeben: $U_{eff} = 230$V. Für andere Kurvenformen ist der Scheitelfaktor verschieden, zum Beispiel $k_s = 1$ für Rechteckspannungen, $k_s = \sqrt{3}$ für Dreieck- oder Sägezahnschwingungen (Angaben jeweils für Spannungen, die symmetrisch zu null sind).

1.1.2 Das Ohm'sche Gesetz

Das Verhältnis zwischen Potentialunterschied (Spannung) an den Anschlüssen eines beliebigen Elementes und dem Strom definiert den Widerstand $\tilde{Z}$:

$$\tilde{Z} := \frac{\tilde{U}}{\tilde{I}} \tag{1.7}$$

Das Ohm'schen Gesetz besagt dabei, dass $\tilde{Z}$ eine *zeitunabhängige* komplexe Grösse ist. Sie heisst die *Impedanz*. Man schreibt

$$\tilde{Z} = R + iX, \qquad \tilde{Z} = Z_0 e^{i\varphi} \tag{1.8}$$

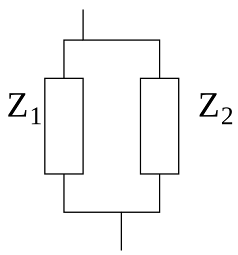
wobei R der *Wirkwiderstand*, X der *Blindwiderstand* und Z_0 der *Scheinwiderstand* genannt werden. Das Inverse der Impedanz $1/\tilde{Z}$ heisst auch der Scheinleitwert oder die *Admittanz.*

Betrachten wir eine harmonische Komponente wie in 1.1 und 1.2, erhalten wir:

$$\tilde{Z} = \frac{U_0 \cdot e^{i(\omega t + \varphi_u)}}{I_0 \cdot e^{i(\omega t + \varphi_I)}} = \frac{U_0}{I_0} \cdot e^{i\varphi} \qquad \text{mit } \varphi = \varphi_u - \varphi_I \tag{1.9}$$

Die komplexe Phase der Impedanz stellt also gerade die Phasendifferenz zwischen Strom und Spannung dar. Da $R = Z_0 \cos\varphi$ werden vor allem bei Leistungsgeräten wie Motoren usw. oft Z_0 (oder U_{eff} und I_{eff}) zusammen mit dem *Kosinusphi* (z.B. $\cos\varphi = 0.7$) spezifiziert. Diese Angaben machen natürlich nur für eine harmonische Spannung mit bestimmter Frequenz Sinn.

Für komplexe Widerstände gelten die gleichen Regeln für Parallel- und Serieschaltung wie bei reellen Widerständen: Bei der Serieschaltung addieren sich die Widerstände (oder die Inversen der Leitwerte).



In der *Parallelschaltung* addieren sich die Leitwerte und somit die Inversen der Widerstände, es gilt also:

$$\widetilde{Z_{tot}} = \frac{\widetilde{Z_1} \cdot \widetilde{Z_2}}{\widetilde{Z_1} + \widetilde{Z_2}} =: \widetilde{Z_1} \| \widetilde{Z_2} \tag{1.10}$$

womit wir die Operation $\|$ definieren, die sehr oft benötigt wird.

(Im folgenden lassen wir die Tilden auf den komplexen Grössen im allgemeinen weg, meinen aber bei Strom, Spannung und Widerstand stets die komplexen Darstellungen.)

1.1.3 Reelle Widerstände, Induktivitäten, Kapazitäten

Alle passiven linearen elektrischen Netzwerke können aus den drei Grundzweipolen reeller Widerstand, Induktivität und Kapazität aufgebaut werden.

Relle Widerstände (auch ohm'sche Widerstände genannt) erzeugen keine Phasendifferenzen zwischen Spannung und Strom, es gilt

$$Z_R = \frac{U_0}{I_0} = R. \tag{1.11}$$

Für *Induktivitäten* gilt

$$U_L = L \cdot \frac{\partial I}{\partial t} \tag{1.12}$$

Bei einem zeitabhängigen Strom $I = I_0 e^{i\omega t}$ entsteht eine magnetisch induzierte Spannung. Durch die Ableitung erscheint ein Faktor $i\omega$, die Impedanz Z_L und deren Betrag, der Scheinwiderstand Z_{0L}, betragen demnach

$$Z_L = \frac{U_L}{I} = i \cdot \omega L, \qquad Z_{0L} = \omega L \tag{1.13}$$

Die Impedanz Z_L ist eine positiv imaginäre Zahl, der Phasenwinkel beträgt demnach $90°$. "Der Strom eilt der Spannung um $90°$ nach".

Für *Kapazitäten* gilt

$$U_C = \frac{1}{C} \cdot Q, \qquad Q = \int I(t) \cdot dt \tag{1.14}$$

Mit einem Strom $I = I_0 e^{i\omega t}$ wird der Kondensator geladen. Durch die Integration des Stromes erscheint ein Faktor $1/i\omega$, die Impedanz Z_C und deren Betrag, der Scheinwiderstand Z_{0C}, betragen

$$Z_C = \frac{U_C}{I} = \frac{1}{i\omega C}, \qquad Z_{0C} = \frac{1}{\omega C} \tag{1.15}$$

Die Impedanz Z_C ist eine negative imaginäre Zahl ($1/i = -i$). Die Spannung eilt dem Strom um $90°$ nach ($\varphi = -90°$).

Ändert der Strom in einen Kondensator nur langsam, so kann die Gleichung 1.14 vereinfacht werden zu

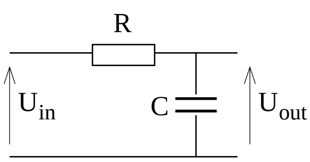
$$U_C = \frac{I \cdot t}{C} \qquad \text{für } I \approx const. \tag{1.16}$$

der sogenannten *Kondensatorgleichung*. Sie ist für viele Abschätzungen sehr nützlich und zeigt, dass sich die Spannung bei konstantem Strom linear mit der Zeit ändert.

Kombinationen von R, L, C

Kombinationen der Grundzweipole können leicht mit dem Ohm'schen Gesetz und den Rechenreglen der komplexen Zahlen berechnet werden. Wir wollen am ersten Beispiel des Tiefpasses auch verschiedene Begriffe des Systemverhaltens kennenlernen. Tief- und Hochpass sind elementare Schaltungen, die überall in der Elektronik sehr häufig vorkommen, deshalb ist die folgende Behandlung etwas ausführlich.

Beispiel Tiefpass:



Bei grossen Frequenzen wird der Scheinwiderstand Z_C der Kapazität immer kleiner. Die Schaltung wirkt als Spannungsteiler, sodass die Ausgangsspannung umso kleiner wird, je höher die Frequenz. Bei tiefen Frequenzen wird Z_C gross, und die Ausgangsspannung wird gleich der Eingangsspannung.

Der gesamte Widerstand des Tiefpasses errechnet sich zu (Serieschaltung!) $Z = R + 1/i\omega C$, der Strom wird nach dem Ohm'schen Gesetz $I = U_{in}/Z$ und die Ausgangsspannung (die Spannung am Kondensator) $U_{out} = I \cdot 1/i\omega C$. Das Verhältnis von Ausgangs- zu Eingangsspannung bezeichnet man auch als *Übertragungsfunktion* G. Stellt man diese für harmonische Eingangsspannungen als Funktion der Frequenz dar, spricht man auch vom (komplexen) *Frequenzgang* $G(\omega)$.

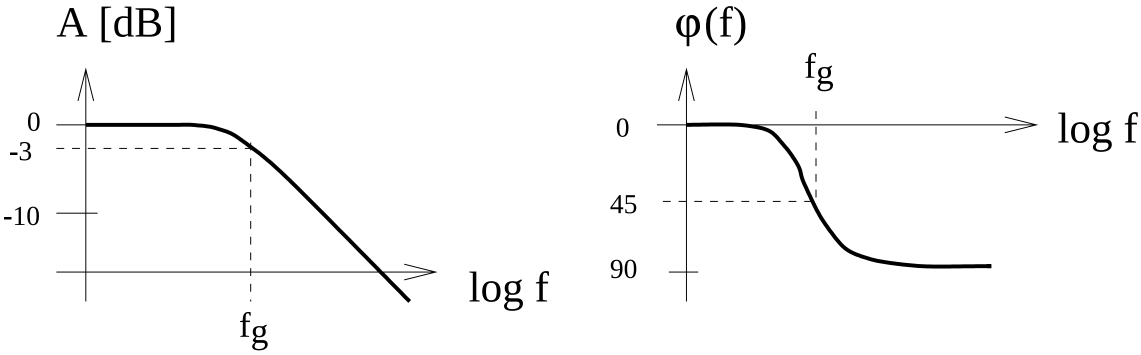
$$G := \frac{U_{out}}{U_{in}} \qquad G(\omega) = \frac{U_{out}(\omega)}{U_{in}(\omega)} \tag{1.17}$$

Für unseren Tiefpass wird also

$$G(\omega) = \frac{I \cdot 1/i\omega C}{Z \cdot I} = \frac{1}{1 + i\omega RC} \tag{1.18}$$

In der Praxis stellt man Betrag und Argument der komplexen Zahl $G(\omega)$ grafisch im sogenannten *Bodediagramm* dar. Der Betrag $|G(\omega)|$ wird auch als *Amplitudengang* bezeichnet und wird im Bodediagramm doppelt logarithmisch als Funktion von ω, das

Argument $\varphi(\omega) = \arg G(\omega)$ einfach logarithmisch dargestellt ($G(\omega) = |G(\omega)| \cdot e^{i\varphi(\omega)}$). Die folgende Skizze zeigt das Bodediagramm für unseren Tiefpass.



Weiter definiert man das *Verstärkungsmass* (englisch attenuation factor) als

$$A(\omega) = 20 \log_{10} |G(\omega)| \tag{1.19}$$

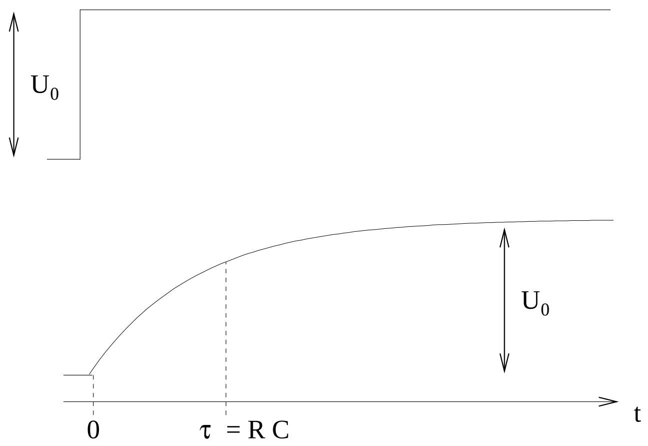
Die Masseinheit für $A(\omega)$ ist das *Dezibel* dB (eigentlich dezi–Bel, das Mass für Leistungsabschwächung, welche proportional dem Quadrat des Verstärkungsmass ist, deshalb der Faktor 20 statt 10).

In unserem Tiefpass bezeichnet man die Frequenz

$$f_g := \frac{\omega_g}{2\pi} := \frac{1}{2\pi RC} \tag{1.20}$$

als *Grenzfrequenz*. Bei derselben ist der Amplitudengang auf $1/\sqrt{2}$ gefallen, das Verstärkungsmass beträgt -3dB. Für Frequenzen unter f_g und für Gleichstrom befinden wir uns im Durchlassbereich, $U_{out} \approx U_{in}$. Für Frequenzen oberhalb der Grenzfrequenz nimmt der Amplitudengang linear mit $1/\omega$ (oder 6 dB pro Oktave) ab, man spricht dann von einem Tiefpass erster Ordnung. Bei der Grenzfrequenz ist die Phase um 45° verschoben, für $\omega \gg \omega_g$ um 90° . Man spricht deshalb von einem Integrator (aus $U_{in} = \cos \omega t$ wird $U_{out} = \frac{1}{\omega} \sin \omega t$).

Oft ist diese Beschreibung eines Systems im *Frequenzraum* der Anwendung nicht besonders gut angepasst. Vor allem für Geräte, die einzelne Ereignisse zum Beispiel Ladungspulse von physikalischen Sensoren verarbeiten sollen, eignet sich die Beschreibung des Verhaltens im *Zeitraum* besser. Dafür definiert man eine *Testfunktion*, zum Beispiel eine Deltafunktion (Impuls), oder der Praxis wesentlich näher stehend eine Stufenfunktion (Sprung) und studiert (meist mit Hilfe von Laplacetransformationen, siehe Kapitel 3.4) die Reaktion des Systems darauf. Man kann zeigen, dass sowohl die *Impulsantwort* als auch die *Sprungantwort*, das Verhalten des Systems unter allen Bedingungen vollständig beschreibt, sofern es sich um ein lineares und zeitinvariantes System (LTI System) handelt (siehe 3.3).



Die Antwort auf einen Spung der Höhe U_0 unseres Tiefpasses ist offensichtlich gegeben durch

$$U_{out} = U_0 \cdot (1 - e^{-\frac{t}{RC}}) \quad (1.21)$$

wie in der nebenstehenden Skizze dargestellt.

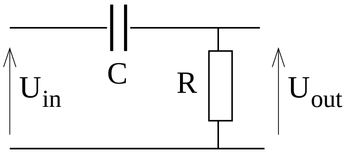
In der Elektronik muss man häufig endliche Anstiegszeiten von Spannungen beschreiben, die eigentlich Sprungfunktionen wären. Man gibt dabei meistens die 10-90 Zeit an, also die Zeit, die es braucht, bis das Signal von 10% bis auf 90% seines endgültigen Wertes gestiegen ist. Man berechnet leicht und merke sich, dass bei einem Tiefpass die Anstiegszeit

$$t_a \approx 2.2\tau \approx \frac{1}{3f_g} \quad (1.22)$$

beträgt. Zum Beispiel benötigen wir für die Beobachtung von Signalen mit einer Anstiegszeit von 1 ns einen Oszillographen mit einer Grenzfrequenz von mindestens 330 MHz. 12.10.98

Der Tiefpass eignet sich ebenfalls als gleitender *Mittelwertbildner* mit der charakteristischen Zeit $\tau = RC$. Schnellere Änderungen werden im Kondensator aufintegriert, lansamere Änderungen werden übertragen. Man kann also Tiefpässe für die Glättung von unruhigen oder mit höheren Frequenzen verrauschten Signalen einsetzen.

Beispiel Hochpass

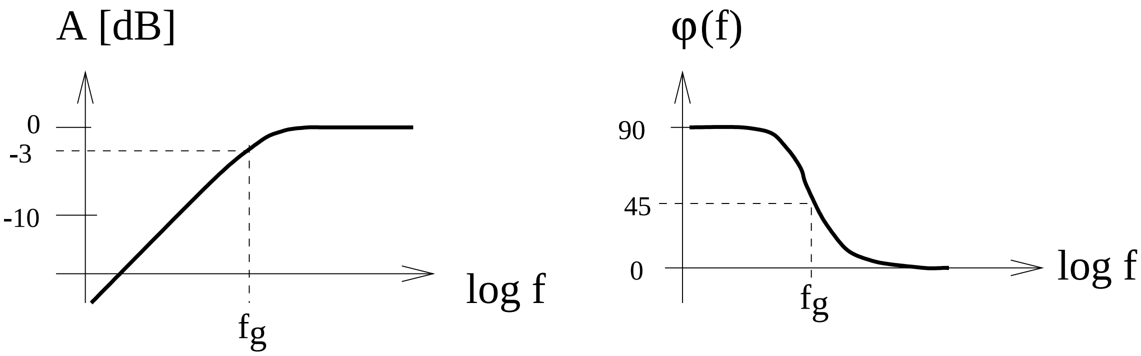


Bei kleinen Frequenzen wird die Impedanz der Kapazität immer grösser, sodass die Ausgangsspannung kleiner wird. Bei grossen Frequenzen wird die Impedanz klein, sodass die Ausgangsspannung gleich der Eingangsspannung wird

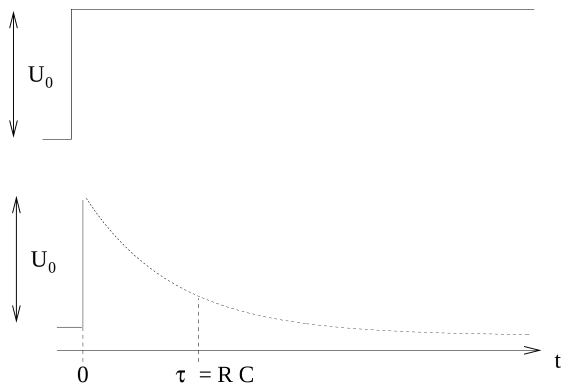
Die Gesamtimpedanz des *Hochpass* ist gleich wie beim Tiefpass $Z = R+1/i\omega C$, nach dem Ohm'schen Gesetz wird $I = U_{in}/Z$. Die Ausgangsspannung ist aber nun einfach $U_{out} = I \cdot R$, sodass der Frequenzgang der Schaltung

$$G(\omega) = \frac{U_{out}}{U_{in}} = \frac{1}{1 + \frac{1}{i\omega RC}} \quad (1.23)$$

wird. Das Bodediagramm ist gegenüber dem Tiefpass an der Grenzfrequenz gespiegelt:



Für sehr kleine Frequenzen wird also aus $U_{in} = \sin \omega t$ ein $U_{out} = \omega \cos \omega t$, der Hochpass wirkt bei kleinen Frequenzen also differenzierend.



Die Sprungantwort nach einem Hochpass veranschaulicht ebenfalls die differenzierende Eigenschaft. Aus einer Sprungfunktion wird angenähert ein Impuls:

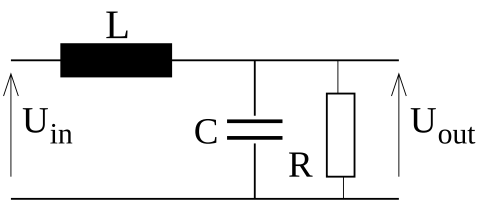
$$U_{out} = U_0 \, e^{-\frac{t}{RC}} \tag{1.24}$$

Der Hochpass kommt insbesondere dort zur Anwendung, wo von einem Signal ein Gleichspannungsanteil entkoppelt werden muss. Zum Beispiel muss ein Proportionalzählrohr mit einer statischen Hochspannung betrieben werden, von der die Verstärkerelektronik mittels eines Hochpasses entkoppelt wird.

Ein anderes Beispiel ist die sogenannte *AC-Kopplung*. Dabei wird mit einem Hochpass, dessen Grenzfrequenz so tief sein muss, dass er die interssanten Signale durchlässt, die Gleichspannungskomponente entkoppelt. Die meisten Oszilloskope haben einen AC – DC Schalter, mit dessen Hilfe sich ein solcher Hochpass in den Signalpfad schalten lässt.

Beachte: Bei all diesen Betrachtungen wurde angenommen, dass der Ausgang der Schaltung unbelastet ist. Wird eine Last, zum Beispiel der Eingangswiderstand eines nachfolgenden Verstärkers oder auch nur ein längeres Kabel angeschlossen, so muss dessen komplexe Impedanz natürlich in die Rechnung eingeschlossen werden. Es ergeben sich dann Situationen wie sie im folgenden Beispiel gerechnet sind:

Beispiel LCR Tiefpass:



Die Anwendung des komplexen ohm'schen Gesetzes liefert die Übertragungsfunktion

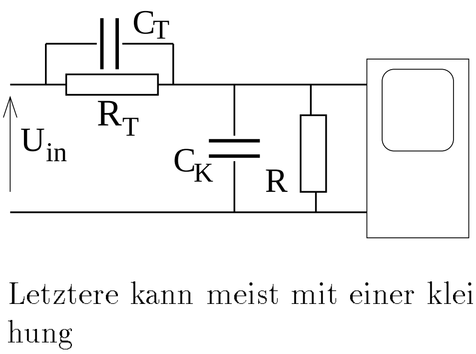
$$G(\omega) = \frac{\frac{1}{i\omega C+\frac{1}{R}}}{i\omega L + \frac{1}{i\omega C+\frac{1}{R}}} = \frac{1}{1-(\frac{\omega}{\omega_0})^2 + \frac{i\omega L}{R}} \quad (1.25)$$

mit $\omega_0 = \sqrt{1/LC}$. Es handelt sich um einen *Serieresonanzkreis* mit Dämpfung R. Bei kleiner Dämpfung (grossem R) divergiert die Übertragungsfunktion bei der Resonanzfrequenz. Wählt man jedoch den Lastwiderstand so, dass $R = \sqrt{L/C}$, dann wird der Amplitudengang bei der Grenzfrequenz ω_0 gerade eins. Bei höheren Frequenzen fällt der Amplitudengang mit $1/\omega^2$ ab, also doppelt so steil wie bei einem RC Tiefpass erster Ordnung.

Die Übertragungungsfunktion hat einen Pol (Nullstelle im Nenner), der im allgemeinen komplex ist. Für $R \rightarrow \inf$ gibt es einen reellen Pol bei $\omega = \omega_0$, d.h. die Schaltung schwingt selbständig. Das ist hier offensichtlich, da es sich um einen Serieresonanzkreis handelt.

Die Diskussion der Existenz und Lage von Polen der Übertragunsgfunktionen spielt in der Entwicklung von Verstärkeren und aktiven Filtern eine grosse Rolle, wir werden das dort ausführlicher diskutieren.

Beispiel Oszilloskop–Tastkopf: Will man ein Signal in einer elektronischen Schaltung oszillographieren, so wird man in der Regel ein Kabel von der Schaltung zum Oszilloskop legen müssen. Damit belastet man aber die zu untersuchende Schaltung kapazitiv, typische Werte eines Koaxkabels sind 100 pF/m. Passive Tastköpfe bestehen aus der folgenden Schaltung:



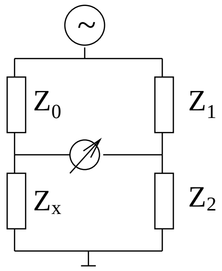
Dabei bedeuten C_K die Kabelkapazität, R der Eingangswiderstande des Oszilloskop (in der Regel 1 MΩ), R_T der Tastkopfwiderstand (in der Regel 9 MΩ) und C_T die Tastkopfkapazität.

Letztere kann meist mit einer kleinen Schraube abgeglichen werden, sodass die Beziehung

$$C_T\ R_T = R\ C_K \quad \text{oder} \quad \frac{C_T}{C_K} = \frac{R}{R_T} \quad (1.26)$$

erfüllt ist. In diesem Fall sind die Argumente der komplexen Widerstände im Tastkopf und am Oszilloskop gleich, es ergibt sich keine Phasenverschiebung. Damit wird das Spannungsteilerverhältnis von der Frequenz unabhängig und die Gesamtbelastung beträgt nur 10 MΩ und einige wenige pF. Ab Frequenzen von etwa $1/2\pi RC \approx 100kHz$ werden allerdings trotzdem die Kondensatoren die Gesamtimpedanz der Messanordnung bestimmen, wobei diese mit $1/\omega$ abnimmt. Für hochfrequentere Anwendungen muss man deshalb aktive Tastköpfe verwenden (FET Eingangsschaltung).

Beispiel Wien–Brücke

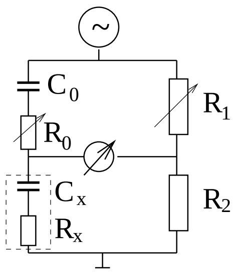


Die Wienbrücke besteht aus vier im allgemeinen komplexen Widerständen, die wie in der nebenstehenden Zeichung angeordnet werden. Ist die in der Mitte gemessene Spannung null, spricht man von einer abgeglichenen Brücke. Dies ist der Fall unter der Bedingung:

$$\frac{Z_0}{Z_x} = \frac{Z_1}{Z_2} \tag{1.27}$$

Dies ist eine komplexe Gleichung, es müssen also sowohl Real- wie auch Imaginärteil der Impedanzen diese Bedingung erfüllen. Im allgemeinen wird die Bedingung nur für eine bestimmte Frequenz erfüllt.

Die Wienbrücke wird zur präzisen Messung von komplexen Impedanzen verwendet. In der Regel verwendet man für Z_1 und Z_2 reelle Widerstände und für Z_0 eine Anordnung von Grundzweipolen die der Ersatzschaltung der unbekannten Impedanz Z_x entspricht.



Die Schaltung zeigt eine *Kapazitätsmessbrücke*, die es erlaubt, sowohl die Kapazität als auch der reelle Anteil der Impedanz eines Kondensators zu bestimmen. Die Amplitude im gemessenen Differenzsignal wird nur null, wenn beide Abgleichwiderstände richtig eingestellt sind:

$$\frac{\frac{1}{i\omega C_0} + R_0}{\frac{1}{i\omega C_x} + R_x} = \frac{R_1}{R_2} \tag{1.28}$$

Die Abgleichbedingung ist nur dann für alle Frequenzen erfüllt, wenn ausserdem die beiden Zeitkonstanten gleich ($R_xC_x = R_0C_0$) sind. Dann wird:

$$C_x = C_0\frac{R_1}{R_2} \qquad \text{und} \qquad R_x = R_0\frac{R_2}{R_1} \tag{1.29}$$

Falls sich in der Messung keine frequenzunabhängige Abgleichbedingung finden lässt, deutet das auf eine unvollständige oder falsche Ersatzschaltung hin.

1.1.4 Bauformen

Leider können ideale R, L und C Elemente nicht gebaut werden. Neben Toleranzen in ihren Widerstandswerten müssen wir auch in Kauf nehmen, dass reale Bauelemente nie rein reell oder imaginär sind. So haben Anschlüsse aller Bauelemente eine Induktivität und einen reellen, endlichen Widerstandswert. Ausserdem gibt es immer zwischen den Anschlüssen eine Kapazität.

Reale Bauelemente werden deshalb durch ein *Ersatzschaltbild* beschrieben, das aus idealen R, L und C Elementen besteht, und dessen Schaltung das Verhalten des realen Bauelementes hinreichend genau beschreibt.

Grundsätzlich unterschiedet man Bauelemente mit Drahtanschlüssen für normale Platinenmontage und *SMD* (surface mounted devices) für die SMT (Surface mounted technology), die nur Anschlussflächen besitzen, und dann auf direkt auf die Oberfläche von Leiterplatten oder Keramiksubstanzen montiert werden. Letztere Technik eignet sich vor allem viel besser für automatische Bestückungs- und Lötverfahren. Die SMT löste einen weiteren Miniaturisierungsschub aus. Sie hat günstigere Hochfrequenzeigenschaften (kleinere Abmessungen, kleinere Anschlussinduktivitäten und Streukapazitäten), wegen der Automatisierung kleinere Bestückungsfehlerraten und kleinere Herstellungskosten.

Für den Laborbetrieb muss man allerdings in Kauf nehmen, dass es viel mühsamer ist, kleine Modifikationen vorzunehmen. Deshalb ist gerade bei Einzelstücken im Experimentierbetrieb die konventionelle Technik manchmal zu bevorzugen.

Für die SMT muss beachtet werden:

- Wärmeproduktion kritischer $\Rightarrow$ verlustarme Techniken verwenden!
- Grössere Bauteile haben manchmal geometrische Wärmeausdehungskoeffiziten, die verschieden sind vom Platinenmaterial
- Im Gegensatz zur herkömmlichen Technik muss man in der Mehrlagenplatine in den inneren Ebenen keine Rücksicht auf die Anschlüsse der Bauteile nehmen, die Platinen können leicht auf beiden Seiten bestückt werden.

Die SMD's werden vor dem Löten mit Klebern fixiert. Dann wird das bereits sich auf der Platine befindliche Lot in einem heissen Dampf erhitzt (typisch 215° C) und so in ca. 10 bis 30 sec. verlötet.

Handlöten geht auch, erfordert aber schon eine ruhige Hand: Lötkolben auf 300° C stellen, Leiterbahn erhitzen nicht Bauelement, warten bis Zinn auf das Bauelement geflossen ist, nicht laňger als 3 Sekunden, sonst werden die Bauelemente zerstört.

Bauformen von Widerständen

(siehe auch [Nühr98], Kapitel 3)

Für die Auswahl von Widerständen für eine bestimmte Anwendung sind neben ihrem Widerstandswert folgende Faktoren relevant:

- Belastbarkeit (maximale Leistung): Typische Werte sind 1/8, 1/4 oder 1/2 Watt, sie sind direkt proportional zur Bau grösse. SMD Widerstände sind in der Regel mit 1/4 Watt belastbar.
- Maximalspannung
- Herstellungsgenauigkeit. Typische Werte sind $\pm 5\%$ im Normalfall. Präzisionswiderstände können bis $\pm 0.5\%$ gehen.
- Rauschen. Neben dem konstruktionsunabhängigen thermischen Rauschen (siehe 1.6) gibt es auch das Stromrauschen (1/f Rauschen), das vom Material und Aufbau des Bauelementes abhängt.
- Temperaturstabilität. Man gibt den Temperaturkoeffizient in ppm/K an.
- Montagetechnik

Der Widerstandswert und die Toleranzen werden mittels Farbringen nach dem Regenbogenschema auf die Widerstände aufgebracht. SMD Widerstände werden nach dem "WWP" Code (Ziffer, Ziffer, Exponent) bezeichnet: z.B. 471 = 470 Ω. Die geometrische Grösse wird mit einer vierstelligen Zahl bezeichnet, wobei die ersten zwei Ziffern Länge und die zweiten zwei Ziffern die Breite in Einheiten von 1/100 inch angeben. Natürlich gibt es auch eine DIN Norm in mm, ausserdem ist dort die Reihenfolge umgekehrt...

Widerstandswerte werden in den sogenannte *E-Reihen* mit ungefähr logarithmischen Abständen produziert. Am häufigsten ist die Reihe E24, das heisst es gibt 24 verschiedenen Werte pro Dekade (100, 110, 120, 130, 150, 160, 180 usw.).

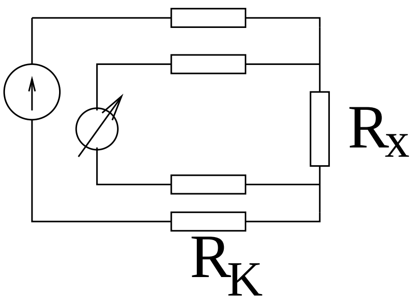
Die Widerstandsschicht kann aus Kohle, Cr/Ni, Gold, Platin oder Metalloxid (SnO_2) bestehen. Kohle hat einen negativen widerstandswertabhängigen, Metalle einen positiven Temperaturkoeffizienten, der vorallem bei Platin sehr genau konstant gehaten werden kann. Deshalb werden Platinwiderstände als Temperaturfühler eingesetzt.

Kohlewiderstände halten die höchsten kurzzeitigen Spannungsspitzen und es gibt sie auch für sehr hohe Werte (Standardwerte bis 22 MΩ). Sie haben eine relativ schlechte Fertigungstoleranz, dafür aber meist ein besseres Hochfrequenzverhalten (unter 200Ω

induktiv, darüber kapazitiv). Trotzdem sind die Effekte noch gross, z.B. hat ein typischer 10 MΩ Widerstand bei 100 MHz noch eine Impedanz von 10kΩ ! Deshalb und auch wegen den Streukapazitäten müssen Hochfrequenzanwendungen immer niederohmig ausgelegt werden. 14.10.98

Nickel-Chrom Drahtwiderstände sind hochpräzis (bis 0.05 %) und sehr temperaturstabil (bis 5 ppm/K) aber nicht besonders klein. Durch geschicktes Wickeln kann man auch die Indutivität in Grenzen halten.

Beispiel 4–Draht–Messung:



Die sogenannte *Kelvinschaltung* ist im nebenstehenden Bild dargestellt. Die Widerstände der Kabel R_K verfälschen bei einer gewöhnlichen Widerstandsmessung den Messwert. Führt man nun den Strom aus einer Konstantstromquelle in einer separaten Leitung dem Widerstand zu, und misst dann auf einer anderen Leitung den Spannungsabfall an R_x, wird das Resultat unabhängig von den Widerständen R_K.

Nichtlineare Widerstände: Im weiteren gibt es drei verschiedene Klassen von *nichtlinearen Widerständen*, nämlich *PTC*, *NTC* und *VDR*. NTC haben einen negativen Temperaturkoeffizient von etwa 2 bis 6%/° K. PTC haben positive Temperaturkoeffizienten im Bereich von 5 bis 70%/° K. NTC und PTC werden auch unter dem Begriff *Thermistoren* zusammengefasst und werden als Temperatursensoren für Regler aller Art verwendet. PTC's im Stromversorgungskreis eignen sich auch direkt als thermischer Überlastschutz: Wird das Gerät zu heiss, nimmt die Stromzufuhr automatisch ab.

VDR (Voltage dependant resistor) ändern ihren Widerstand in Funktion der Spannung und bestehen meist aus Zinkoxid. Der Widerstand nimmt exponentiell mit der Spannung ab und die Reaktionszeit dafür beträgt nur wenige ns, sodass sich VDR sehr gut als Spannungsstossbegrenzer (z.B. wenn Induktivitäten im Schaltkreis vorhanden sind) und Überspannungsschutz eignen. VDR gibts mit Kniespannungen von etwa 10 bis 700 V.

Bauformen von Kondensatoren

Kondensatoren bestehen im Prinzip aus zwei leitenden Flächen mit Isolationsmaterial dazwischen mit möglichst grosser Dieelektrizitätskonstante ϵ. Als Isolationsmaterial verwendet man Luft, Keramik, Kunststofffolien, Glimmer, Papier und Elektrolyte.

Im Ersatzschaltbild müssen Serieinduktivität des Anschlusses (typisch 1 nH pro mm), Seriewiderstand und der parallele Isolationswiderstand berücksichtigt werden.

Die Kondensatoren werden charakterisiert durch

- Kapazitätswert (von pF bis mF) und Toleranz (1 bis 20%)
- maximale Betriebsspannung, sowohl für Gleichspannung (Nennspannung) als auch für den überlagerten Wechselspannungsanteil (Spitzenspannung).
- Verlustfaktor wegen des Seriewiderstandes ($\tan\delta = \omega RC$), gleichzeitig das Verhältnis von Wirk- zu Blindleistung. Typische Werte gehen von 10^{-3} bis 0.5

Für höchste Frequenzen und kleine Kapazitäten werden Keramikkondensatoren (Typ I, $\epsilon \approx 100$) verwendet. Typische Verlustfaktoren liegen bei $0.3 \cdot 10^{-3}$, der Isolationswiderstand ist in der Grössenordnung von 10^{10} Ω . Für die Induktivität gehen nur die Anschlussdrähte und die Baulänge der Kapazität ein (1 nH/mm), sodass auch für mehrere 100 MHz der Scheinwiderstand noch durch die Kapazität dominiert bleibt (SMD ist noch günstiger!). Diese Kondensatoren haben die höchste Präzision und die beste Stabilität. Es sind Werte bis etwa 1 nF erhältlich.

Keramikkondensatoren Typ II haben ein $\epsilon > 1000$ und sind deshalb weniger präzis und weniger stabil, sie haben auch grössere Verluste. Sie sind im Wertebereich bis 0.2 μF erhältlich

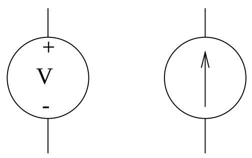
Im Kapazitätsbereich von 1 nF bis 1μF werden bevorzugt Folienkondensatoren eingesetzt, die entweder aus aufgewickelten Isolationsfolien (Kennbuchstabe K für Kunststoff oder P für Papier) bestehen, die mit Metall bedampft sind (vorangestellter Kennbuchstabe M), oder es werden eine Metallfolien und eine Isoationsfolie zusammenaufgewickelt. Ein weiterer Buchstabe gibt die Art des Kunststoffes an. z.B. besteht ein MKS Kondensator aus eine Kunststoffdielektrikum aus Polystyrol mit metallischen Belägen. Es werden Verlustfaktoren bis 10^{-4} erreicht. Die aufgrund der aufgewickelten Folie und den Anschlussleitungen effektiv vorhandene Serieinduktivität führt zu einem Serieresonanzkreis mir Eigenfrequenz im Bereich von 10 bis 100 MHz (kleiner bei grösserer Kapazität).

Elektrolyt- und Tantalelektrolyt Kondenstoren werden für grössere Kapazitätswerte verwendet. Der Isolationswiderstand ist schlechter, gesamthaft werden Verlustfaktoren im Bereich 10^{-2} bis 1 erreicht, die Resonanzfrequenzen liegen in der Grössenordnung 100 kHz.

Elektolytkondensatoren sind polarisiert, d.h. sie funktionieren nur bei *einem* Vorzeichen der Spannung. Einer der beiden Anschlüsse ist deshalb mit einem +, manchmal auch mit einem dicken Strich gekennzeichnet, der den positiven Anschluss bezeichnet. Wird ein solcher Kondensator mit einem falschen Vorzeichen betrieben, sind Kapazität und Verlustfaktor völlig anders, bei zu hoher negativer Spannung pflegt er sich spontan in seine Einzelteile zu zerlegen.

1.2 Strom- und Spannungsquellen

Um einen Stromkreis zu betreiben, braucht man mindestens entweder eine Spannungs- oder eine Stromquelle. Ideale Stromquellen liefern immer den gleichen Strom undabhängig von der an den Anschlüssen auftretetender Spannung. Ideale Spannungsquellen liefern eine konstante Spannung, unabhängig vom fliessenden Strom.



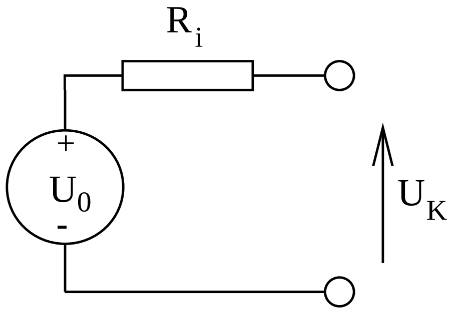
Die symbolische Darstellung einer idealen Spannungs- und Stromquelle

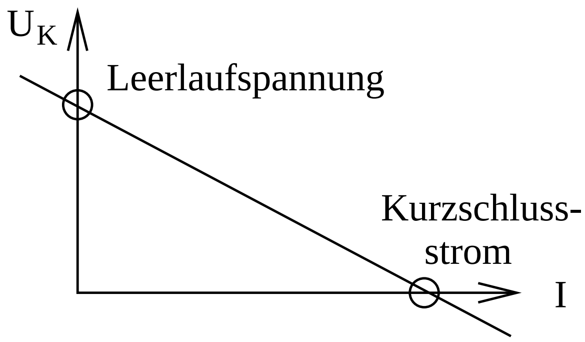
1.2.1 Reale und lineare Strom- und Spannungsquellen

Ideale Strom- und Spannungsquellen existieren nicht, schon deswegen weil sie ein unendliches Energiereservoir darstellen würden.

Die Spannung einer realen Spannungsquelle (z.B. einer Batterie) ist abhängig von dem abgegebenen Strom. Die Spannung an den Anschlüssen, die sogenannte *Klemmenspannung* U_k sinkt mit zunehmendem Strom. Man sagt auch: Die Spannung bricht zusammen, wenn man die Spannungsquelle belastet.

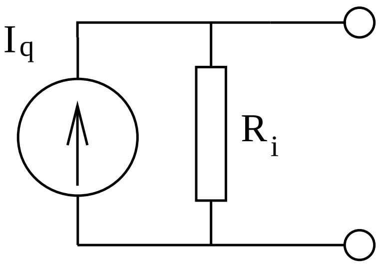
Eine etwas realistischere, aber immer noch nicht reale, sogenannte lineare Spannungsquelle wird durch ein *Ersatzschaltbild* beschrieben, das aus einer ideale Spannungsquelle mit einem (im allgemeinen Fall komplexen) *Innenwiderstand* R_i besteht:

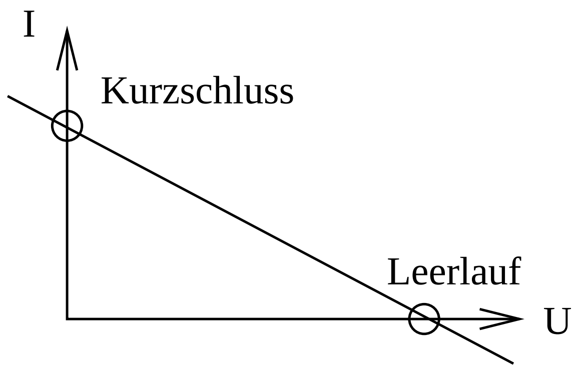




Die Ausgangsspannung U_K nimmt nach dem ohm'schen Gesetz linear mit dem Strom ab: $U_K = U_0 - R_i I$. Den Zusammenhang zwischen Strom und Spannung an einer Quelle nennt man allgemein die *Ausgangskennlinie*, die für unsere Spannungsquelle im rechten Bild skizziert ist. Bei *Kurzschluss*, d.h. bei $U_K = 0$, fliesst der Kurzschlussstrom $I_k = U_0/R_i$. Eine Spannungsquelle nennt man *niederohmig* (resp. *hochohmig*), wenn ihr Innenwiderstand R_i klein (resp. gross) ist. Fliesst kein Strom, spricht man von *Leerlauf*, dann wird $U_K = U_0$.

Lineare Stromquellen haben einen Innenwiderstand, der parallel zur Stromquelle liegt:

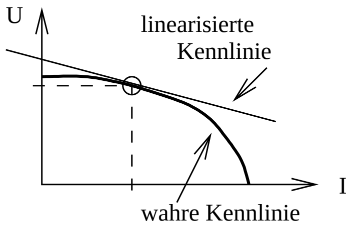




Der Ausgangsstrom nimmt bei höherer Spannung ab, da nun ein immer gröserer Teil des Stromes durch den Innenwiderstand fliesst. Es ergibt sich ebenfalls eine lineare Kennlinie. Im Kurzschluss fliesst der Quellenstrom I_q, im Leerlauf beträgt die Spannung $U = I_q R_i$. Eine gute Stromquelle ist besonders hochohmig.

Da die beiden Kennlinien beide linear sind, kann man deren Ersatzschaltbilder durcheinander ersetzen: Eine Stromquelle I_q mit Innenwiderstand R_i ist identisch zu einer Spannungsquelle mit Leerlaufspannung $U_0 = I_q R_i$ und gleichem (!) Innenwiderstand R_i.

Praktische Spannungs- bzw. Stromquellen haben höchstens in einem eingeschränkten Bereich der Ausgangskennlinie eine lineare Charakteristik. Das Ausgangssignal einer Verstärkerschaltung zum Beispiel kann bei kleiner Last einen relativ kleinen Innenwiderstand haben. Bei sehr grossen Strömen nimmt der Innenwiderstand oft zu, sodass man eine nichtlineare Ausgangskennline bekommt.



Solange jedoch der Strom nur wenig ändert, ist eine lineare Betrachtung hinreichend: Man linearisiert die Kennlinie am Arbeitspunkt, und berechnet damit das Verhalten der Schaltung bei kleinen Variationen. Dieses Verfahren heisst *Kleinsignalanalyse*.

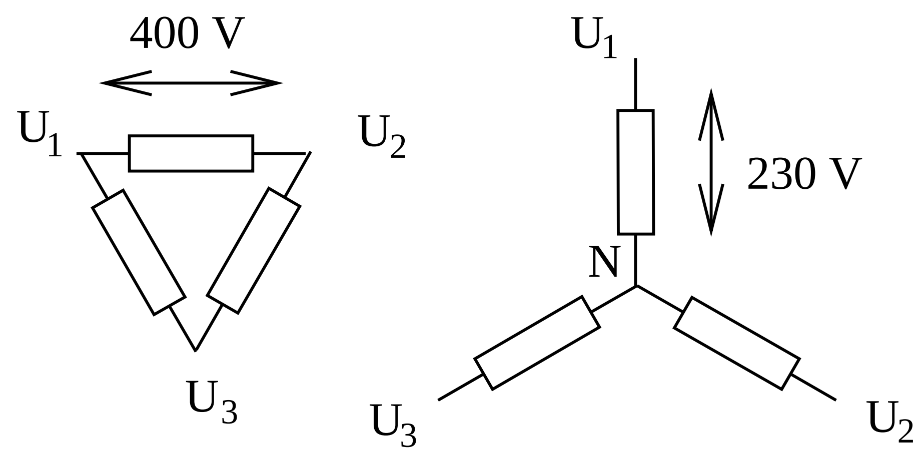
Man sieht, dass der Unterschied zwischen Spannungs- und Stromquellen nicht wohldefiniert ist. Man nennt Quellen mit am Arbeitspunkt eher flachen Kennlinien Spannungsquellen, solche mit steilen Kennlinien Stromquellen.

1.2.2 Drehstrom

Betrachtet man mehrere Wechselstromkreise fester Frequenz, die untereinander eine feste Phasenbeziehung haben, spricht man von *Mehrphasensysteme*. Symmetrische Mehrphasensysteme haben eine gleichmässig verteilte Phasendifferenz über 2π und die gleiche Spannung. Verbindet man jeweils einen Leiter der verschiedenen Systeme miteinander, so erhält man ein verkettetes Mehrphasensystem.

Das internationale Verteilsystem für elektrische Energie ist ein verkettetes symmetrisches Dreiphasensystem und wird auch *Drehstromnetz* genannt. Die drei Wechselstromsysteme haben beim Endverbraucher jeweils eine Effektivspannung von 400 V und sind zueinander um $2\pi/3 = 120°$ phasenverschoben.

Man spricht von Dreieck- (allgemein Polygon-) und Sternschaltungen:



Die Effektivspannungen in der *Dreieckschaltung* betragen 400 V, in der *Sternschaltung* $400/\sqrt{3} = 230$V. Die Anschlusspunkte (*Phasen*) werden mit U_1, U_2, U_3 (früher R, S, T oder U, V, W) bezeichnet. Der symmetrische Nullpunkt ist spannungslos und wird als *Neutralleiter* N (früher M_p oder O) bezeichnet.

Der Neutralleiter ist bei symmetrischer Last auch in der Sternschaltung stromlos. Bei unsymmetrischer Last im Alltag ist das jedoch im allgemeinen nicht der Fall, da man verschiedene 230 V Verbraucher hat. Dadurch entstehen Spannungsschwankungen auf dem Neutralleiter wegen dessen endlichen Leiterwiderstand und er eignet sich deshalb nicht sehr gut als Massepotential.

Deshalb wird in der Regel noch eine sogenannte *Schutzerdung* installiert, bei der man sich bemüht, sie stromlos zu halten, und die gut mit Erdung (Wasserleitungen etc.) verbunden ist. Für störungsfreie ("brummfreie") elektronische Installationen ist es oft angebracht, deren Bezugsmasse mit der Schutzerdung zu verbinden, aber keinesfalls mit dem Neutralleiter!

Der Vorteil des Drehstroms besteht in der Verkettung: Dadurch kann mit nur drei Leitern soviel Energie übertragen werden, wie bei drei unabhängigen Systemen, was sechs Leiter erfordern würde. Durch die Wahl zwischen Stern- oder Dreieckschaltung hat der Verbraucher ausserdem zwei verschiedene Spannungen zur Verfügung. (Historisch gesehen war vor allem die einfache Bauweise von Drehstromasynchronmotoren von ausschlaggebender Bedeutung, heute können die dafür notwendigen Drehfelder elektronisch erzeugt werden). 19.10.98

1.3 Die Kirchhoff'sche Regeln

Gustav Robert Kirchhoff lebte von 1824 bis 1887 und war Professor in Heidelberg von 1854 bis 1876. Die nach ihm benannten Regeln für ein Netzwerk von Zweipolen lauten wie folgt:

1. *Knotenregel* (Kirchhoff's current law, KCL): In einem Knoten ist die Summe aller Ströme null.
2. *Maschenregel* (Kirchhoff's voltage law, KVL): In einem geschlossenen Stromkreis ist die Summe aller Spannungen null.

Das physikalische Prinzip der Knotenregel ist die Ladungserhaltung, das der Maschenregel die Energieerhaltung (im konservativen Feld verschwindet die Potentialdifferenz entlang einer geschlossenen Kurve).

Man braucht in der Praxis Konventionen: Bei einem Knoten gehen positive Ströme vom Knoten weg (ist nicht einheitlich in der Literatur). In der Masche nimmt man positive Spannungsabfälle entlang der positiven Stromrichtung, Gleichspannungsquellen werden von + nach - positiv gerechnet.

Beispiel: Berechne die Spannung U_L an einem Verbraucher einer reellen Spannungsquelle: Der Stromkreis besteht aus einer idealen Spannungsquelle U_0, einem Innenwiderstand R_i und einem Lastwiderstand R_L.

$$\text{KVL:} \quad -U_0 + IR_i + IR_L = 0 \tag{1.30}$$

Daraus erhält man $I = U_0/(R_i + R_L)$ und somit

$$U_L = \frac{R_L}{R_i + R_L} \cdot U_0 \tag{1.31}$$

Man spricht von einem *Spannungsteiler*. Der Stromkreis heisst offen, wenn kein Strom fliesst und man spricht von Kurzschluss, wenn $R_L \ll R_i$. Dann ist die U_L fast null. In diesen beiden Fällen verschwindet die Leistung im Lastwiderstand.

Man spricht von *Leistungsanpassung*, wenn die Leistung im Lastwiderstand maximal wird. Das ist bei $R_L = R_i$ der Fall: $P = U_0^2/4R_L$. Bei Leistungsanpassung wird gleichviel Leistung am Innenwiderstand wie am Lastwiderstand verbraucht. Die Leistungsanpassung spielt ebenfalls bei der Übertragung schneller Signale auf langen Leitungen eine Rolle (siehe 1.5).

Beachte, dass in einem Netz von Zweipolen, die Anwendung nur einer der beiden Regeln genügend Gleichungen liefert, um alle Spannungen und Ströme zu berechnen. Die verschiedenen Lösungsverfahren für die Berechnung eines linearen Netzwerkes werden entsprechend in Knotenpotentialanalyse und Maschenstromanalyse eingeteilt (siehe [Kori98], Seite 32).

1.4 SPICE Simulationen

Das Programm SPICE wurde ursprünglich im Rahmen einer Dissertation an der Universität in Berkley, Ca. entwickelt. In der Zwischenzeit ist es stark ausgebaut worden und kommerziell erhältlich, leider werden die Algorithmen aber nicht mehr offengelegt.

Das Grundprinzip besteht aus der Lösung eines linearen (!) Netzwerkes von m Zweipolen mit Hilfe eines Knotenpotentialverfahrens. Dabei wird jedem Knoten ein Potential U_i zugeordnet. Die Ströme in einem Zweipol mit Impedanz Z_{ik} und Leitwert $Y_{ik} = 1/Z_{ik}$

zwischen den Knoten i und k werden durch $(U_i - U_k) \cdot Y_{ik}$ ausgedrückt. Allfällige Spannungsquellen werden durch ihre äquivalente Stromquellen mit Leerlaufstrom I_q ersetzt (siehe 1.2.1). Die Anwendung der Knotenregel ergibt für jeden der n Knoten des Netzes eine lineare Gleichung. Es entsteht ein System von n inhomogenen linearen Gleichungen

$$f(Y_{ik}) \cdot \underline{U} = \underline{I_q} \tag{1.32}$$

wobei $f(Y_{ik})$ eine Matrix ist, die die m Zweipole enthält, $\underline{U} = (U)_i$ der Vektor der Knotenpotentiale und $\underline{I_q}$ der Vektor der Stromquellen, den die einzelnen Knoten sehen.

Meistens gibt es eine Lösung (siehe lineare Algebra). SPICE berechnet nun diese Lösung, wobei als Randbedingung ein Potential als Masse definiert, und somit null sein muss. Es braucht mindestens eine Stromquelle im System, sonst erhält man nur die Triviallösung $\underline{U_0} = 0$. Ausserdem müssen solche Knoten, die nur kapazitiv mit dem Rest des Systems verbunden sind, mit einer Anfangsbedingung (z.B. mit sehr hohem Widerstand gegen Masse) versehen werden, sonst weigert sich SPICE weiterzurechnen. Diese Rechnung wird als "Bias Calculation" bezeichnet.

Man kann nun auch eine zeitliche Änderung eines Potentiales vorgeben (z.B. Sprungfunktion oder auch beliebige Funktionen der Zeit). SPICE kann dann für das darauffolgende Zeitintervall mit dieser neuen Randbedingung wieder eine Lösung berechnen. Das Verfahren lässt sich fortsetzen und wird "Transientenanalyse" genannt.

Weiter lässt sich mit Hilfe einer Wechselspannungsquelle "VAC" eine Schaltung mit variabler Frequenz speisen, und direkt Bodediagramme für das Übertragungsmass und die Phasenverschiebung erstellen ("AC Analyse").

Man beachte, dass man von linearen Gleichungen ausgeht. Das erfordert also zuerst eine Linearisierung von Kennlinien aller vorkommenden Bauelemente im Arbeitspunkt, wie in 1.2.1 besprochen.

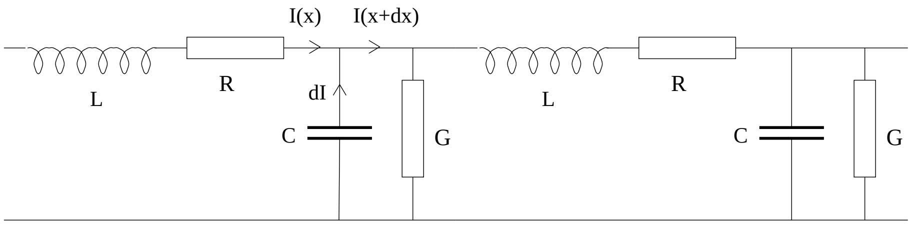
Das Programm existiert unter dem Namen pSpice für Windows und ist leider nicht gerade billig. Eine Demoversion, die auch schon einiges kann, gibt es im Internet (siehe: http://www.microsim.com/ bzw. http://www.orcad.com/). Es ist aber auch in den meisten anderen elektronischen CAD Systemen eingearbeitet.

```
Demonstrationen:
1. Tiefpass erster Ordnung: Eingabe, Bias Calculation
      VAC, AC Analyse
      Transientenanalyse mit Stepfunction,
2. LRC Glied:
      AC Analyse: steilere Funktion
      falsches R (=10*R) -> Serieresonanz.
2. Kapazitiver SPannungsteiler f"ur Sparkdetection
```

1.5 Endliche Leitungen

(Vollständiger Formalismus siehe z.B. [Putz71], Seite 325)

Leitungen dienen der Übertragung elektrischer Signale und Energie. Leitungen endlicher Länge mit angelegter Spannung und in denen ein Strom fliesst, wechselwirken elektromagnetisch mit sich selbst und der Umgebung. Die *Maxwellgleichungen* beschreiben dieses System vollständig. Die Eigenschaften von Leitungen können jedoch auch vollständig mit einer geeigneten Ersatzschaltung aus den besprochenen Grundzweipolen R, L, C beschrieben werden:



Die Abbildung zeigt eine solche, die in der Praxis die meisten Fälle beschreibt. Mit L, R, C und G sind jeweils die Grössen pro Leitungslänge gemeint. R ist der ohm'sche Widerstand der Signalleitung, C die Kapazität zwischen Signal- und Masseleitung, L die Selbstinduktivität der Leitung und G der Leitwert der Isolation (=1/Verlustwiderstand). Betrachte man das Leitungsstück dx indem der Strom $I(x,t)$ fliesst , so ergibt sich ein Spannungsabfall $U(x,t)$ über L und R nach dem ohm'schen Gesetz:

$$-\frac{\partial U}{\partial x} = (R + i\omega L)\ I \tag{1.33}$$

Mit der Knotenregel bekommen wir ausserdem:

$$-\frac{\partial I}{\partial x} = (G + i\omega C)\ U \tag{1.34}$$

Differenzieren wir die erste Gleichung nach x und eliminieren wir den Term $\frac{\partial I}{\partial x}$. Es bleibt eine Differentialgleichung für die Spannung $U(x)$ übrig, die sogenannte Telegraphengleichung:

$$\frac{\partial^2 U}{\partial x^2} = (R + i\omega L)\ (G + i\omega C)\ U \tag{1.35}$$

(analog für den Strom). Als Lösung erhalten wir gedämpfte Wellen entlang dem Leiter:

$$U(x,t) = U_0 e^{i\omega t - \gamma x}, \qquad \text{mit} \quad \gamma = \alpha + i\beta = \sqrt{(R + i\omega L)(G + i\omega C)} \tag{1.36}$$

γ heisst die komplexe Dämfungskonstante. Ihr Realteil α beschreibt die Dämpfung, der Imaginärteil $\beta = \frac{2\pi}{\lambda}$ die Wellenzahl. Die Phasengeschwindigkeit unserer Welle ist:

$$v = \frac{\omega}{\beta} \tag{1.37}$$

Für eine verlustfreie Leitung mit $R = 0$ und $G = 0$ wird $\beta_0 = \omega\sqrt{LC}$ und somit $v = v_0 := 1/\sqrt{LC}$ konstant.

Man nennt das Verhältnis von Spannung zu Strom in der Leitung analog dem ohm'schen Gesetz den *Wellenwiderstand* oder die *Wellenimpedanz* Z. Durch Einsetzen der Lösung für U direkt in 1.33 erhält man

$$Z = \frac{U}{I} = \sqrt{\frac{R + i\omega L}{G + i\omega C}} \tag{1.38}$$

also im allgemeinen eine komplexe Zahl. Eine verlustfreie Leitung hat die Impedanz $Z = Z_0 := \sqrt{L/C}$.

Beispiele:

Koaxkabel bestehen aus einer konzentrischen Anordnung von Innenleiter (Durchmesser d), Isolation und Aussenleiter (Innendurchmesser D). Die Induktivität und Kapazitäte pro Längeneinheit berechnen sich zu

$$L = \frac{\mu}{2\pi} \cdot \ln\frac{D}{d}, \qquad C = \frac{2\pi\epsilon}{\ln\frac{D}{d}} \tag{1.39}$$

und betragen numerisch typisch zu 250 nH/m bzw. 100 pF/m. Die Fortpflanzungsgeschwindigkeit wird

$$v = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\mu\epsilon}} \tag{1.40}$$

und hängt damit nur vom Isolationsmaterial und nicht von der Geometrie ab. Typische Werte sind 60% der Lichtgeschwindigkeit, also 5ns/m. Ganz schnelle Kabel bekommt man mit $\epsilon = 1$, sogenannte Luftkabel, wo nur mit minimalem Material der Innenleiter an Ort gehalten wird. – Die Impedanz wird schliesslich

$$Z = \sqrt{\frac{L}{C}} = \sqrt{\frac{\mu}{\epsilon}} \cdot \frac{1}{2\pi} \cdot \ln\frac{D}{d} \tag{1.41}$$

hier kommt es also auf das Isolationsmaterial und auf das Durchmesserverhältnis an. Die meisten Koaxkabel haben $Z = 50\ \Omega$, es ist aber auch 35, 75 und 95 Ω gebräuchlich.

Twisted Pair Kabel zu deutsch *verdrillte Leitungen* genannt, bestehen aus zwei symmetrischen isolierten Litzen (Druchmesser d), die im Abstand a gehalten werden und verdrillt werden. Dadurch wird das Einkoppeln von Störsignalen (Übersprechen) stark reduziert. Induktivität und Kapazität pro Längeneinheit werden

$$L = \frac{\mu}{\pi} \ln \frac{2a}{d}, \qquad C = \frac{\pi \epsilon}{\ln \frac{2a}{d}} \tag{1.42}$$

Numerische Werte sind etwa 600 nH/m und 50 pF/m, die Impedanzen liegen etwa bei 120 Ω . Dabei ist zu beachten, dass durch die schlechter definierte Geometrie die Impedanz ziemlich variieren kann.

Koaxkabel haben zwar die bessseren Eigenschaften, sind aber teurer und brauchen mehr Platz. Twisted pair Kabel eignen sich vor allem für die Übertragung digitaler Signale.

Striplines bestehen aus Kupferbahnen (Breite b, Dicke$d$), die auf einer gedruckten Schaltung der Dicke D aufgebracht sind. Auf deren Rückseite befindet sich eine durchgehende Kupferlage, die die Masseschicht darstellt. Die Impedanz dieser Anordnung beträgt:

$$Z = \frac{75}{\sqrt{\epsilon}} \cdot \ln \frac{6D}{0.75b + d} + \frac{0.075b}{D} \quad [\Omega\ , \mathrm{mm}] \tag{1.43}$$

Typische Kupferkaschierungen haben eine Dicke von $d = 35\ \mu$m, Leiterplatten meist eine solche von ca. 1.5 mm (oder 0.4 mm bei einem 4–lagen Print). Der einzig freie Parameter um eine bestimmte Impedanz zu erreichen, bleibt deshalb die Breite b. Für $Z = 100\Omega$ wird benötigt man eine Leiterbahnenbreite von ca. 0.6 mm für 1.5 mm dicke Leiterplatten (0.15 mm für 0.4 mm Dicke). Die Ausbreitungsgeschwindigkeit beträgt auch hier ca. 5 ns/m. Sie wird aber um Faktoren grösser, wenn an den Leiterbahnen IC's angeschlossen werden: Deren Eingangskapazität verlangsamt die Signalausbreitung, da $v = 1/\sqrt{LC}$!

Dämpfung:

Für den Fall einer sogenannten *Widerstandsdämpfung* ($R > 0$, aber $G = 0$) erhalten wir aus Gleichung 1.36 ($v_0 = 1/\sqrt{LC}$):

$$\beta = \frac{\omega}{v_0} \sqrt{\frac{1}{2}\,(1 + \sqrt{1 + \frac{R^2}{\omega^2 L^2}})} \tag{1.44}$$

$$\alpha = \frac{\omega RC}{2\beta} \tag{1.45}$$

Damit wird für $\omega \approx \omega_0 = R/L$ die Wellenzahl β nicht mehr genau proportional zu ω, sondern etwas grösser. Die Phasengeschwindigkeit ω/β somit frequenzabhängig kleiner, d.h. wir haben *Dispersion*, die Signale werden verzerrt übertragen. Ebenfalls ist dann Z frequenzabhängig (siehe 1.38) und wird für kleine Frequenzen grösser.

Für kleine Frequenzen, bzw. grosse Drahtwiderstände ($\omega \ll \omega_0$) wird $\beta = \sqrt{\omega RC/2}$ und damit $\alpha = \beta$. Das heisst die Dämpfung wird mit der Wurzel der Frequenz schlimmer. Für grosse Frequenzen bzw. kleine Drahtwiderstände wird $\beta = \omega/v_0$, $Z = Z_0$ und demnach $\alpha = R/2Z_0$. Die Dispersion verschwindet also für hohe Frequenzen.

Der *Skineffekt* bewirkt aber, dass die Stromdichte an der Oberfläche eines metallischen Körpers gegen innen exponentiell abnimmt (siehe zum Beispiel [Jack90]. Die mittlere Eindringtiefe des Stromes ist

$$\delta = \sqrt{\frac{2\rho}{\omega\mu}} \tag{1.46}$$

(ρ Leitfähigkeit). Typische Werte sind für Kupfer bei 1 MHz 70μm, bei 1 GHz 2 μm. Deshalb nimmt der effektive Kabelwiderstand mit $\sqrt{\omega}$ zu, und entsprechend hat man auch bei grossen Frequenzen noch eine $\sqrt{\omega}$ – Abhängigkeit der Dämpfung. Um die Oberfläche zu vergrössern, nimmt man für HF – Anwendungen deshalb manchmal Kupferfolien oder mehrere Litzen. 26.10.98

Zahlenbeispiel: Koaxkabel RG 178: $Z = 50\Omega$, $R = 0.45$ Ω /m, $L = 240$ nH/m, $C = 95$ pF/m. Bei 2 MHz und 100 m Länge wird das Signal um einen Drittel reduziert. Bei 200 MHz überleben nur gerade noch 1% des Signals nach 100 m. Bessere Kabel müssen dicker sein. Die kritische Frequenz wird etwa $\omega_0 = 2$ MHz. In diesem Bereich gibt es Dispersion, die kleineren Frequenzen kommen verspätet an (Tiefpassverhalten). Die Dispersion wird offensichtlich signifikant, wenn $\omega > \omega_g = (Rl \cdot Cl)^{-1}$, der Tiefpassgrenzfrequenz. Für z.B. 20 MHz ist das bei einer Kabellänge von ca. $l = 30$m der Fall. Beachte, dass die Tiefpassgrenzfrequenz mit der Kabellänge im Quadrat kleiner wird, die Übertragungsqualität eines Kabels wird demnach mit dem Quadrat der Länge schlechter.

Falls gerade $R/L = G/C$ ist, spricht man von einer *verzerrungsfreien Leitung*, da dann die Dispersion gerade verschwindet (v =const) und die Impedanz und die Dämpfung frequenzunabhängig werden.

Die Sache ist allerdings wie immer in der Praxis komplizierter, da es auch Verluste wegen frequenzabhängiger Polarisation des Dielektrikums gibt. Bei ganz hohen Frequenzen schliesslich werden die Hohlleitermode wichtig, das ist aber erst in der Grössenordnugn 10 GHz der Fall. Für den praktischen Einsatz eines Kabels muss man die Abschwächung und Phasenverzerrung messen, bzw. die numerischen Herstellerangaben zu Rate ziehen.

Abschluss:

Ändert sich die Geometrie der Leitung an einer bestimmten Stelle, so ändern sich auch Ausbreitungsgeschwindigkeit und Impedanz. Mit dem gleichen Formalismus wie für ein Wellenpaket am endlichen Potentialtopf in der Quantenmechanik oder in der Optik kann man die reflektierte und die transmittierte Welle berechnen (siehe zum Beispiel [Hin96], Seite 21-24).

Ein spezieller Fall ist das Ende einer Leitung. Es sei mit einem Widerstand R abgeschlossen. Ist $R = Z$, so ist der Widerstand “angepasst”, die Energie der ankommenden Welle wird im Widerstand vernichtet. Mit dem erwähnten Formalismus erhält man den *Reflexionsfaktor* für die Amplitude der Welle:

$$r = \frac{R - Z}{R + Z} \tag{1.47}$$

Offene Leitungen, also grosses R, führen zu Reflexionen von Signalen mit gleichem Vorzeichen, kleine Widerstände (Kurzschlüsse) führen zu Reflexionen mit umgekehrtem Vorzeichen. Anwendung: Messung von Leitungslängen und Lokalisierung von Impedanzinhomogenitäten.

Insbesondere Jede Art von geometrischer Inhomogenität entlang der Leitung (Steckverbindungen, Lötstellen usw.) führt zu Reflexionen und Verlusten. Hochfrequenzleitungen müssen deshalb sehr sorgfältig gebaut werden.

Abschluss bei mehreren Quellen oder Verbraucher

Sollen an einer Leitung mehrere Verbraucher angeschlossen werden, oder sollen verschiedene Quellen die gleiche Leitung treiben (Bussystem), dann gelten folgende Regeln:

- Die Leitung muss eindimensional sein, allfällige Abzweigungen für Anschlüsse müssen kurz gegenüber der Wellenlänge sein.
- Beide Enden der Leitung müssen mit Z_0 abgeschlossen sein.
- Alle Verbraucher (ausser an den Enden der Leitung) müssen eine grosse Eingangsimpedanz haben $R_i \gg Z_0$ und möglichst reell sein. Kapazitive Anschlüsse verlangsamen die Signalausbreitung.
- Die Dämpfung der Leitung soll klein sein $R \ll \omega L$, anderenfalls sind Zwischenverstärker vorzusehen.

Kurze Leitungen:

All diese Betrachtungen der Wellenausbreitungen sind nur sinnvoll, wenn die Wellenlänge klein ist gegenüber der Leitungslänge. Kurze Leitungen, also wenn $l \ll \lambda$, verhalten sich wie eine Kapazität, wenn der Ausgang unbelastet ist. Schliesst man die kurze Leitung am Ausgang kurz, sieht man am Eingang die Induktivität.

Will man ein Signal aus einer Spannungsquelle mit endlicher Innenwiderstand R_i über eine solche relativ kurze Leitung übertragen, stellt das ganze also ein Tiefpass mit $\tau = R_i \cdot C \cdot l$ dar.

Sogenannte $\lambda/4$ – *Leitungen* haben eine Länge $l = \lambda/4$ und verhalten sich etwa so wie eine gedackte Orgelpfeife. Offene $\lambda/4$ – Leitungen stellen einen Serieresonanzkreis ($|Z| = 0$, aber) für diese Wellenlänge dar. Schliesst man eine solche Leitung kurz, erhält man einen Parallelresonanzkreis.

```
Demo langes Koax-Kabel:

1. Signalverz"ogerung mit kurzem Puls
2. durchstimmbarer Abschlusswiderstand: pos und neg Reflexionen
3. Bananenstecker Verbindung nix gut.
4. Kabel gut abschliessen: ANstiegsflanke zeigt trotzdem Dispersion.
5. Kabel Kurzschluss.
6. L"ange messen mit offenem Kabel
7. l"angerer Puls mit Reflexionen: Treppenstufen
8. Kleine Frequenz mit R_i=10kOhm: Tiefpass
9. Lambda Viertel Leitung mit Resonanz.
```

1.6 Physik des Rauschens

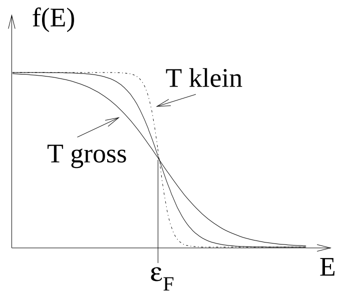
Einschub
21.10.98

Kapitel 2

Halbleiter – Bauelemente

2.1 Halbleiter und ihre Dotierung

In einem isolierten Atom sind die diskreten Energieeigenzustände der Hüllenelektronen bei sehr niedriger Temperatur von unten her aufgefüllt. Mit Fermienergie ϵ_F bezeichnet man die Grenze ab welcher die Energieniveaus leer sind. Erhöht man die Temperatur T werden einzelne Elektronen mit einer gewissen Wahrscheinlichkeit auf höheren Niveaus sich aufhalten.



Die *Fermifunktion* beschreibt dann die Belegungsdichte $f(E)$ der Zustände mit Energie E in Funktion der Temperatur T:

$$f(E) = \frac{1}{1 + e^{\frac{E-\epsilon_F}{kT}}} \tag{2.1}$$

Man merke sich, dass $kT \approx 1/40$ eV bei Zimmertemperatur.

Bringen wir nun viele Atome in einer regelmässigen Anordnung (Gitter) unter, dann kombinieren die diskreten Energieniveaus der einzelnen Atome in gemeinsame *Energiebänder*. Noch immer wird die Belegungsdichte von der Fermifunktion beschrieben. Von besonderem Interesse sind die beiden Bänder unmittelbar in der Nähe der Fermienergie: Das *Valenzband* befindet sich unmittelbar unterhalb der Fermienergie, das *Leitungsband* unmittelbar darüber. Der Abstand zwischen Valenz- und Leitungsband heisst *band gap energy* E_G.

Bei $T = 0$ befinden sich alle Elektronen im Valenzband (bzw. in noch tieferen Niveaus). Bei erhöhter Temperatur werden einzelne Elektronen in das Leitungsband angehoben. Diese n_e *Leitungselektronen* und die dadurch entstehenden n_p *Löcher* im Valenzband sind die Träger der Leitfähigkeit des Materials.

Bei reinen Halbleitern spricht man von *intrinsischer Leitung*: Es gibt gleich viele Leitungselektronen wie Löcher: $n_p = n_e =: n_i$. Aus der Fermifunktion lässt sich leicht ableiten, dass

$$n_i \sim e^{-\frac{E_G}{2kT}} \tag{2.2}$$

.

Die folgende Tabelle zeigt einige Beispiele, wobei für die Halbleiter auch die intrinsische Leitungsdichte bei 300 K angegeben ist. Diese ist absolut sehr klein im Vergleich zu der Zahl der Atome, die für ein reines Siliziumgitter zum Beispiel etwa $5 \cdot 10^{22}$ cm^{-3} beträgt.

		E_G [eV]	n_i [cm^{-3}]
Isolatoren	SiO_2	8	
	Diamant	5	
Halbleiter	GaAs	1.42	$2 \cdot 10^6$
	Si	1.12	$1 \cdot 10^{10}$
	Ge	0.66	$2 \cdot 10^{13}$
Metalle		$\simeq 0$	

28.10.98

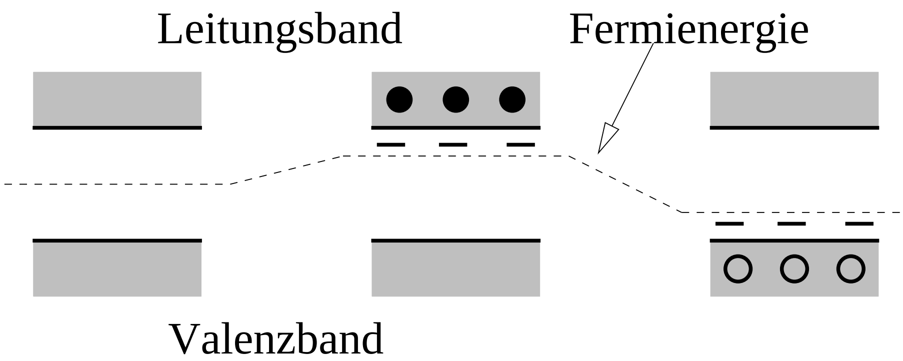
Der überwiegende Teil aller Halbleiterelektronik wird heute aus Silizium hergestellt. Der wesentliche Trick besteht nun in der *Dotierung* (englisch *Doping*). Dabei werden minimal kleine Mengen von 3 oder 5 wertigen Atomen als Verunreinigung in das Gitter der 4 wertigen Silizium Atome eingebracht.

Wird ein 5-wertiges Atom an einer Gitterstelle eingebracht, die eigentlich für ein 4-wertiges Atom gedacht ist, gibt es einen zusätzlichen Elektronenzustand, der allerdings sehr schwach gebunden ist. Sein Energieniveau liegt nahe am Leitungsband. Man spricht von einem *Donator*, wofür typischerweise Phosphor oder Arsen verwendet wird.

Ebenso ergibt sich bei einer 3-wertigen Verunreinigung einen zuätzlichen Lochzustand, dessen Energieniveau nahe am Valenzband liegt und das leicht ein Elektron aus dem Valenzband binden kann. Typische *Akzeptoren* sind Bor, Aluminium oder Gallium.

Die Bindungsenergien für die zusätzlichen durch die Dotierung mit Bor oder Phosphor entstehenden Elektronen- oder Lochzustände betragen nur 0.045 eV, sind also in der Grössenordnung von kT bei Raumtemperatur. Deshalb befinden sich im normalen Betrieb die meisten der zusätzlichen Elektronen im Leitungsband (ebenso die Löcher im Valenzband). Die Dotierungsstärke bestimmt die Leitfähigkeit.

Man beachte aber, dass der Leiter bei der Dotierung natürlich insgesamt neutral bleibt. Die durch die Ionenimplementierung zusätzliche positive Ladung wird durch einen entsprechenden Zustrom an Elektronen ausgeglichen.



Das Bild links zeigt die Lage der Fermienergie für undotierte Halbleiter. Bei n-Dotierung ergeben sich zusätzliche Elektronenzustände wie in der mittleren Skizze gezeigt. Die Fermienergie verschiebt sich deshalb nach oben. Analog dazu im rechten Bild die Situation für Lochzustände. Bei 300 K sind die Elektronen und Löcher wie skizziert fast alle im Leitungs- bzw. Valenzband. n-Dotierung führt also zu zusätzlicher Elektronleitung, p-Dotierung zu zusätzlicher Lochleitung.

Einen Zusammenhang zwischen der Verschiebung der Fermienergie und der Dichte der Leitungselektronen erhält man aus der Fermiverteilung und einigen Annahmen über die effektive Lage und Dichte der Dotierungszustände für Halbleiter im Gleichgewicht (für eine Herleitung siehe zum Beispiel [Pier96], Seite 49-53):

$$n_e = n_i \cdot e^{(\epsilon_F - \epsilon_i)/kT} \tag{2.3}$$

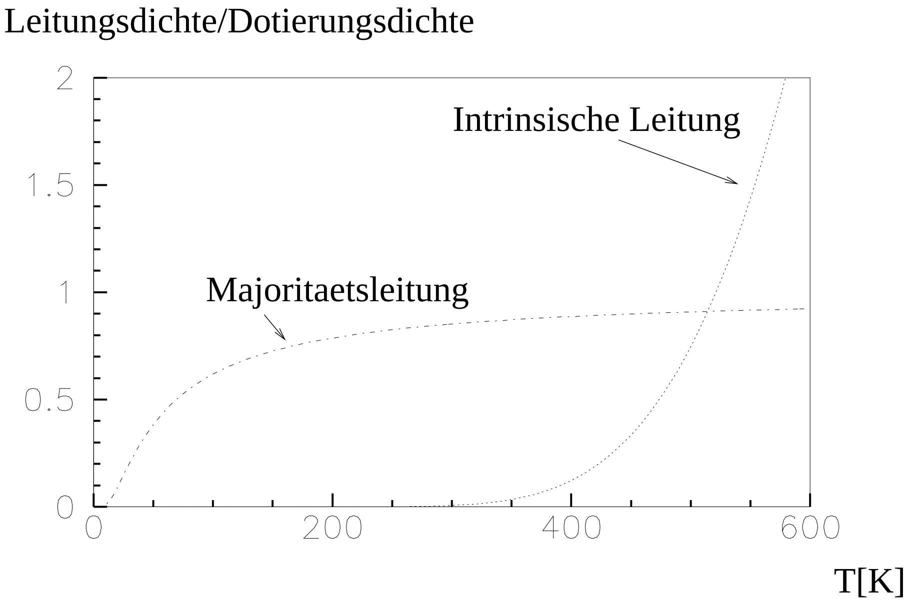
$$n_p = n_i \cdot e^{(\epsilon_i - \epsilon_F)/kT} \tag{2.4}$$

Dabei sind n_i und ϵ_i die Leitungsdichte und die Fermienergie im intrinsichen Halbleiter, ϵ_F die durch die Dotierung verschobene Fermienergie. Die Gleichungen gelten nur, falls der Abstand zwischen ϵ_F und dem Leitungs- bzw. Valenzband $\geq 3kT \approx 75$mV ist, andernfalls nennt man den Halbleiter degeneriert.

Daraus folgt unmittelbar auch das sogenannte np Produkt:

$$n_e \cdot n_p = n_i^2 \tag{2.5}$$

Dotiert man also einen Halbleiter mit Donatoren, die zu einer Leitungselektronendichte n_e führen, kann die Lochdichte mit der obigen Formel abgeschätzt werden. Die Temperaturabhängigkeit von n_i ist durch 2.2 gegeben. Falls die Dotierung gross gegenüber der intrinsichen Leitfähigkeit ist (typische praktische Werte in Silizium sind 10^{15}cm^{-3}), werden schon bei relativ kleinen Temperaturen alle zugehörigen Elektronen sich im Leitungsband befinden, die Leitungsdichte wird gleich der Dotierungsdichte.



Die *Majoritätsleitung* (die durch die Dotierung hervorgerufene Leitung) ist ab einer gewissen Temperatur also konstant, wenn alle dotierten Ladungsträger sich im Leitungsband befinden: $n_e = N_D$.

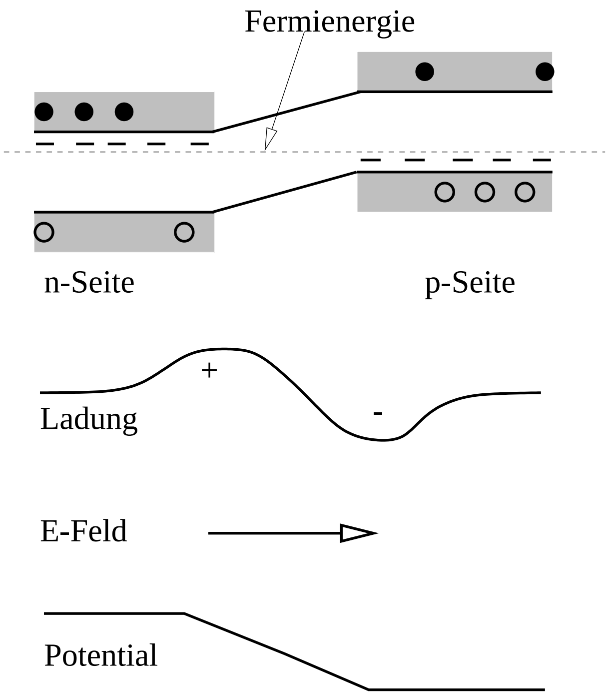
Die Leitung der nicht dotierten Ladungsträger heisst *Minoritätsleitung*. Aus Gleichung 2.5 folgt für die Minoritätsleitungsdichte für einen n-dotierten Halbleiter $n_p = n_i^2/n_D$, sie hängt mit Gleichung 2.2 also stark von der Temperatur ab. Bei grossen Temperaturen dominiert dann die intrinsische Leitung, das Halbleiterbauelement wird in der Regel unbrauchbar. (Die Figur ist für eine Phosphor - Dotierung in Silizium mit $10^{15}\mathrm{cm}^{-3}$ berechnet, siehe [Pier96], Seite 66).

2.2 Die *pn* Grenzschicht

Durch Ionenimplementationstechniken ist es heute möglich auch sehr kleine Strukturen von verschieden dotierten Bereichen in Silizium herzustellen. Der im englischen *pn junction* genannte Übergangsbereich zwischen p- und n-dotierten Halbleiterbereichen stellt das zentrale Element der Halbleitertechnik dar.

Denken wir uns, dass zwei vorher gegeinander isolierte p und n-Bereiche in perfektem Kontakt gebracht werden. Da es auf der p-Seite viel mehr Löcher gibt als auf der n-Seite, werden diese durch Diffusion in den n-Bereich wandern. Ebenso diffundieren Elektronen in den p-Bereich. Das hat zur Folge, dass im Grenzbereich eine Ladung entsteht, die durch die dotierten Gitterstellen gebildet wird, denen jetzt die Ladungsträger fehlen.

Diese Ladung hat nun ein elektrisches Feld und somit ein Potential zur Folge, das *Diffusionspotential*, das der Diffusion entgegenwirkt. Es bildet sich also ein Gleichgewicht zwischen Diffusion und Potential. Die Energiebänder arrangieren sich so, dass die Fermienergie gerade eine Konstante ist:

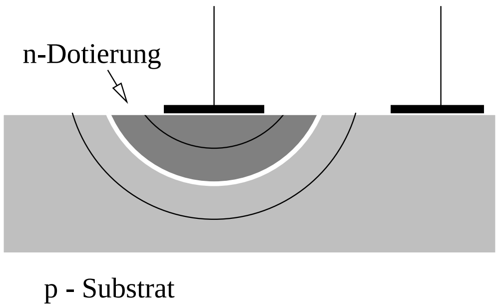


Der Grenzbereich hat eine wesentlich reduzierte Leitungsdichte, man spricht von der *Verarmungszone* (*depletion region*) oder von der hochohmigen *Sperrschicht*.

Das Diffusionspotential kann aus der erwähnten Gleichgewichtsbedingung berechnet werden ([Pier96], Seite 203), unter Verwendung der sogenannten Einstein – Beziehung, die die Diffusionskonstante, die Ladungsträgerbeweglichkeit und die Temperatur miteinander verknüpft. Man erhält:

$$U_d = \frac{kT}{e} \ln(\frac{N_A N_D}{n_i^2}) \tag{2.6}$$

Dabei sind N_A und N_D die Akzeptor- und Donatordichten auf der p- und n- Seite des Überganges. Für eine typische Siliziumsperrschicht mit 10^{15} cm^{-3} Dotierung auf beiden Seiten erhält man eine Diffusionsspannung von 0.6 V. U_d ist offensichtlich temperaturabhängig mit einem Koeffizienten von etwa 2mV/K.



Bei der Herstellung verwendet man zum Beispiel ein p dotiertes Silizium - Substrat (typischerweise Scheiben mit 6" = 150 mm Durchmesser). Durch Ionenimplementation von Phosphor wird ein stark n-dotierter Bereich geschaffen. An der Grenze bildet sich die depletion region

Die ideale *pn* Kennlinie

(siehe Skizze in [Pier96], Seite 236)

Wird nun an diesen Halbleiterübergang eine externe Spannung angelegt, dann verschieben sich die Potentiale der Energieniveaus und die Fermienergie der beiden Seiten gegeneinander:

Legt man eine negative Spannung (*reverse bias*, Minus an der p-Seite) an, so wird der Potentialwall vergrössert. Es können nur die Minoritätsträger einen Strom bewirken. Kommt zum Beispiel ein Elektron aus der p-Seite durch Diffusion zufällig zu nahe an den Potentialwall, wird es durch das in der Sperrschicht herrschende elektrostatische Feld auf die n-Seite gezogen, es fliesst ein kleiner Strom I_r. Die Ursache dieses Strom ist also Leitungsträger – Diffusion ausserhalb der Sperrschicht, er hängt deshalb nicht von der angelegten Spannung ab. Wegen $n_e = n_i^2/N_A$ ist dieser sogenannte *Sperrstrom* jedoch stark temperaturabhängig $I_r \sim e^{-E_G/kT}$ (vgl. Gleichung 2.2).

Legt man eine positive Spannung (*forward bias*, Plus an der p-Seite) an, so wird der Potentialwall abgebaut. Da gemäss der Fermifunktion die Leitungsladungsdichte im Valenz- und Leitungsband exponentiell mit dem Abstand von der Fermienergie abnimmt, wird die Wahrscheinlichkeit, dass ein Ladungsträger den Potentialwall überwindet, exponentiell von der angelegten Spanung abhängen.

Mit diesen qualitativen Überlegungen können wir die ideale Kennlinie der *pn* Grenzschicht wie folgt ansetzen:

$$I = I_r \cdot (e^{\frac{eU}{kT}} - 1) \tag{2.7}$$

Diese Gleichung wird auch ideale *Diodengleichung* genannt, der damit beschriebene Strom nennt man entsprechend seinem Ursprung den *Diffusionsstrom*. Diese einseitige Leitung der Grenzschicht ist die Grundlage der Funktion aller *junction* Halbleiterelemente.

Sperrschichtdicke

(siehe [Pier96], Seite 210-214)

Da die Majoritätsträger durch die Polung der anliegenden Sperrspannung von der Sperrschicht zurückgezogen werden, vergrössert sich die Breite der Sperrschicht mit wachsender Spannung.

Vorerst erhalten wir aus der Bedingung, dass in der Sperrschicht die Gesamtladung verschwinden muss (Gauss'scher Satz), eine Beziehung für die relative Ausdehnung der Sperrschicht in den p- und n- Bereich x_p und x_n, wenn man als Randbedingung die Donator- und Akzeptor-Dotierungsdichten N_D und N_A verwendet:

$$eN_A x_p = eN_D x_n \quad \text{somit} \quad \frac{N_A}{N_D} = \frac{x_n}{x_p} \tag{2.8}$$

Die Sperrschicht dehnt sich also umso mehr in einen Bereich aus, je kleiner dessen Dotierung im Vergleich zur anderen Seite ist.

Aus der eindimensionalen Poissongleichung

$$\frac{\partial E}{\partial x} = -\frac{\rho}{\epsilon\epsilon_0} \tag{2.9}$$

und der Randbedingung, dass $\int E dx$ über die gesamte Sperrschicht die Spannung $U - U_d$ ergeben muss, erhalten wir ferner:

$$U - U_d = -\frac{eN_D}{2\epsilon\epsilon_0} x_n^2 - \frac{eN_A}{2\epsilon\epsilon_0} x_p^2 \tag{2.10}$$

Diese zwei Gleichungen mit zwei Unbekannten sind lösbar, für die gesamte Sperrschichtdicke erhalten wir dann

$$d = x_p + x_n = \sqrt{\frac{2\epsilon\epsilon_0}{e}(\frac{1}{N_A} + \frac{1}{N_D}) \cdot (U_d - U)} \tag{2.11}$$

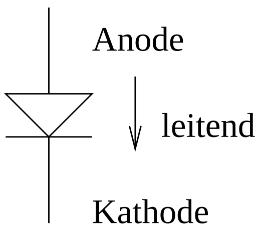
Die Dicke hängt also von der Wurzel der angelegten Spannung an und die kleinere Dotierungskonzentration bestimmt hauptsächlich die Gesamtdicke.

Die **Kapazität** der Sperrschicht hängt von deren Fläche und der Dicke d ab. Für schnelle Schaltkreise ist eine kleine Kapazität erforderlich, weil man dann beim Schalten weniger Ladung abfliessen lassen muss. Dafür braucht man kleine Ausdehnungen und grosse Dicken, man muss also kleine Dotierungsdichten wählen. Da $d \sim \sqrt{U}$ kann man die Kapazität mit der Sperrspannung steuern, was in den *Kapazitätsdioden* seine Anwendung findet. Anwendung: VCO (Voltage controlled oscillator).

Ebenfalls eine dicke Sperrschicht und somit kleine Dotierungsdichten braucht man für hohe **Spannungsfestigkeit** im Sperrbetrieb.

Im Gegensatz dazu bewirkt eine hohe Dotierungsdichte zwar eine kleine Sperrschichtdicke, wegen Gleichung 2.5 aber eine kleinere Minoritätsladungsdichte, und somit einen kleineren Sperrstrom. 2.11.98

2.3 Dioden und ihre Anwendungen



Das einfachste Halbleiterelement mit einem einzigen *pn* Übergang heisst *Diode*. Halbleiterdioden werden für einen weiten Anwendungbereich hergestellt. Die p-Seite heisst auch Anode, die n-Seite Kathode, letztere ist meist mit einem Ring gekennzeichnet, also die Seite von der kein Strom fliessen kann.

Man spricht auch von Gleichrichterdioden, ihre charakteristischen elektrischen Eigenschaften sind

- maximale Sperrspannung
- Sperrstrom bei Raumtemperatur
- Kapazität und Schaltgeschwindigkeit
- maximaler Vorwärtsstrom

und natürlich ihre geometrische Grösse und die Anschlusstechnik.

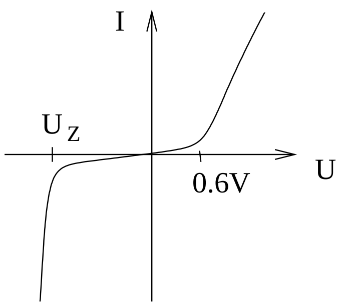
2.3.1 Reale Kennlinie

Schaut man sich die Kennlinie einer realen Diode genauer an, stellt man sofort fest, dass es erhebliche Abweichungen gibt zum Diffusionsstrom, der durch die ideale Diodengleichung 2.7 beschrieben wird.

Der Sperrstrom ist höher als von der Dotierung erwartet und hängt doch etwas von der Spannung ab. Das liegt an der thermischen Erzeugung von Elektron-Lochpaaren im Sperrbereich. Der Effekt wird umso grösser, je dicker die Sperrschicht (mehr Volumen) und ist somit proportional zu $\sqrt{U}$.

Bei kleinen Spannungen im Vorwärtsbereich gibt es den umgekehrten Effekt. Neben dem idealen Diodenstrom gibt es auch solche Ladungen, die zwar den Potentialwall noch nicht überschreiten können, aber innerhalb der Sperrschicht zufällig auf ein Loch mit dem gleichen Problem treffen. Die so entstehende Rekombinationswahrscheinlichkeit führt zu einem effektiv erhöhten Strom, der Effekt ist propotional zu $e^{eU/2kT}$.

Die beiden erwähnten Prozesse werden unter dem Begriff *Recombination–Generation–Current* zusammengefasst, man spricht vom I_{R-G}. (ausführliche Diskussion in [Pier96], Seite 270)

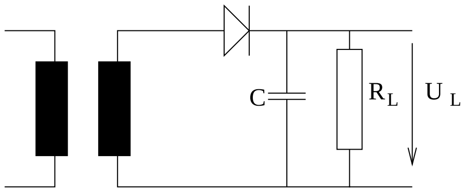


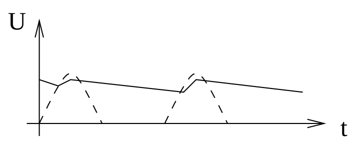
Bei sehr grossen Vorwärtsströmen beginnt sich das Halbleitermaterial wie ein ohm'scher Widerstand zu verhalten, der Anstieg der Kennlinie wird dadurch zunehmend linear statt exponentiell.
Bei grossen Sperrspannungen schliesslich bricht die Sperrschicht zusammen, die Diode wird schnell leitend, man spricht vom Zenerbereich (siehe Abschnitt 2.3.4). Zusammenfassend resultiert die Diodenkennline im nebenstehenden Bild.

2.3.2 Dioden als Gleichrichter

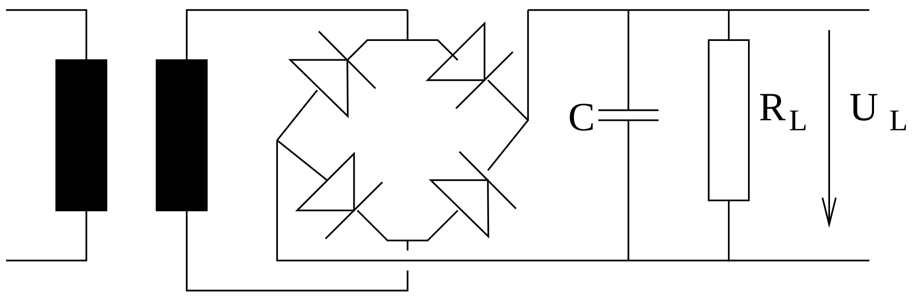
Die Leistungsversorgung aller unserer Geräte erfolgt über das Wechselspannungsnetz. Elektronikanwendungen benötigen aber Gleichspannung, weshalb Gleichrichter in die Netzteile eingebaut werden müssen. Man unterscheidet zwischen Vollwellen- und Halbwellengleichrichter.

Beim Halbwellengleichrichter wird nur eine der Halbwellen der Wechselspannung verwendet. Ein Kondensator wird auf die Spitzenspannung der Wechselspannung abzüglich der Durchlassspannung der Diode aufgeladen, und sorgt dafür, dass auch in der Pause weiter Strom fliesst.





Beim Vollwellengleichrichter in der Brücken- oder Grätzschaltung (unten) werden beide Halbwellen benützt, was die Pausen zwischen den Spannungsspitzen kürzer macht und demnach nur einen halb so grossen Kondensator benötigt. Allerdings gehen nun zweimal die Diodendurchlassspannung verloren. Beim symmetrischen Vollwellengleichrichter werden zwei symmetrische Spannungen erzeugt, indem man bei gleicher Diodenbeschaltung dem Transformator noch eine Mittenanzapfung als Masse verpasst, und dann von beiden Gleichspannungsanschlüssen je einen Kondensator gegen Masse schaltet.



2.3.3 p.i.n. Dioden

Zur Erzeugung hoher Sperrspannungen baut man *pin Dioden*, die aus einer Schichtfolge p.i.n. bestehen, das heisst aus einer p-dotierten, einer nicht dotierten, intrinsischen und einer n-dotierten Schicht bestehen. Die beiden dotierten Schichten sind sehr stark dotiert, sodass sich grosse Ströme erreichen lassen, aber trotzdem hohe Sperrspannungen, wegen dem grossen Abstand durch den i Anteil. Man erreicht Ströme bis 1 kA und einige kV Sperrspannung.

PIN Dioden werden auch als Detektoren für geladene Teilchen in der Kern- und Teilchenphysik eingesetzt: In den Sperrrichtung betriebenen Dioden erzeugt die Ionisationsladung einen zusätzlichen Strom, der detektiert werden kann.

2.3.4 Zenerdioden

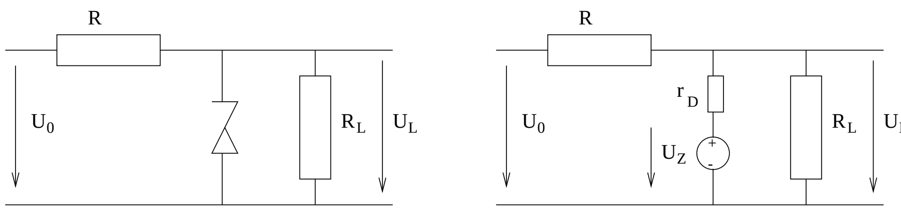
Bei sehr grossen Sperrspannungen bricht die Diode zusammen, der Strom nimmt plötzlich sehr grosse Werte an. Dafür sind zwei unterschiedliche Prozesse relevant. Der sogenannte *Zenerstrom* kommt durch den quantenmechanischen Tunneleffekt zustande, die *Zenerspannung* hängt damit von der Sperrschichtdicke ab. Der Effekt ist nur bei sehr hoher Dotierung relevant und kann mit verschiedenen Dotierungsstärken von etwa 2 bis 8 V variiert werden. Typische Dotierungen sind 10^{17}cm^{-3}, was eine Sperrschichtdicke von etwa 0.01 μm ergibt. Man beobachtet experimentell, dass die Zenerspannung mit zunehmender Temperatur abnimmt.

Bei grösseren Sperrspannungen von schwächer dotierten Dioden wird der *Lawinendurchbruch* relevant. Wenn das elektrische Feld in der Sperrschicht sehr stark wird, erfahren die Elektronen, die den Dunkelstrom erzeugen, eine hohe Beschleunigung. Wird deren Energie bis zum nächsten Stoss mit dem Gitter hoch genug, können sie neue Elektron-Lochpaare erzeugen, welche wiederum stark beschleunigt werden. Dieser Lawineneffekt führt beim Überschreiten einer gewissen kritischen Spannung zu einem plötzlich stark ansteigenden Strom. Man beobachtet einen Zunahme der Durchbruchspannung mit der Temperatur.

Man spricht – eigentlich unkorrekt – bei allen Dioden, die speziell für den Betrieb an der Durchbruchspannung vorgesehen sind, von *Zenerdioden*.

Etwa in der Gegend von 6 V sind die Effekte von Zener- und Lawinendurchbruch gleich gross. Wegen den unterschiedlichen Vorzeichen der Temperaturkoeffizienten, erhält man in diesem Bereich sehr temperaturstabile Spannungsreferenzen, von typischerweise $\frac{\partial U_Z}{\partial T}\frac{1}{U_Z} \approx 10^{-4}/\mathrm{K}$.

Eine solche Spannungsstabilisierung besteht aus einem Vorwiderstand der die Schwankungen dU_0 der unstabilen Spannung U_0 im Verhältnis zum differentiellen Widerstand r_D der Zenerdiode reduziert. Der differentielle Widerstand ist definiert als die Steigung der Kennlinie im Arbeitspunkt der Zenerdiode, die wir im Ersatzschaltbild als eine Serieschaltung von einer idealen Spannungsquelle und einem Widerstand r_D auffassen:



Das Bild zeigt also links die Stabilisierungsschaltung mit einer Last und rechts das zugehörige Ersatzschaltbild für die Kleinsignalanalyse. Mit dem Knotenpotentialverfahren erhalten wir die Gleichung

$$\frac{U_0 - U_L}{R} = \frac{U_L - U_Z}{r_D} + \frac{U_L}{R_L} \tag{2.12}$$

Daraus berechnet man U_L in Funktion von U_0. Eine kleiner Änderung von U_0 bewirkt eine solche von U_L, man erhält schliesslich für eine Dimensionierung von $R/r_D \gg U_0/U_Z$

$$\frac{\mathrm{d}U_L}{U_L} = \frac{r_D\, U_0}{R\, U_Z} \cdot \frac{\mathrm{d}U_0}{U_0} \tag{2.13}$$

Das Ergebnis ist scheinbar unabhängig von R_L was sehr angenehm ist. Natürlich darf aber der Strom durch R_L nur so gross werden, dass der Arbeitspunkt der Zenerdiode an der gewünschten Stelle bleibt. Dieser Sachverhalt geht in der Herleitung wegen der Linearisierung der Kennlinie unter! Ausserdem erzeugt der volle Strom im Widerstand R eine Verlustleistung von $R \cdot I^2$, die bei grossen Spannungsunterschieden oder grossen Strömen den Einsatz dieser Schaltung verbietet. Man spricht von *analogen Netzteilen*, oder *längsgeregelten Netzteilen*. Die Stabilisierung wird in der Regel mit Verstärkerschaltungen verbessert, der Widerstand wird dabei durch einen Transistor als regelbaren Widerstand ersetzt.

Es sind komplete Schaltungen als IC's erhältlich mit der Identifikationsnummer 78xx für positive und 79xx für negative Spannungen, wobei xx den Wert der geregelten Ausgangsspannung bedeutet. Diese Bausteine haben 3 Anschlüsse (Eingang, Ausgang, Masse), der Eingang muss mindestens 2 Volt über dem Ausgang liegen. Sie sind thermisch und elektrisch überlastgeschützt.

Solche längsgeregelte Netzteile werden heute praktisch nur noch für kleine Ströme ($<$ 1A) eingesetzt, da sonst der Kühlaufwand zu gross wird.

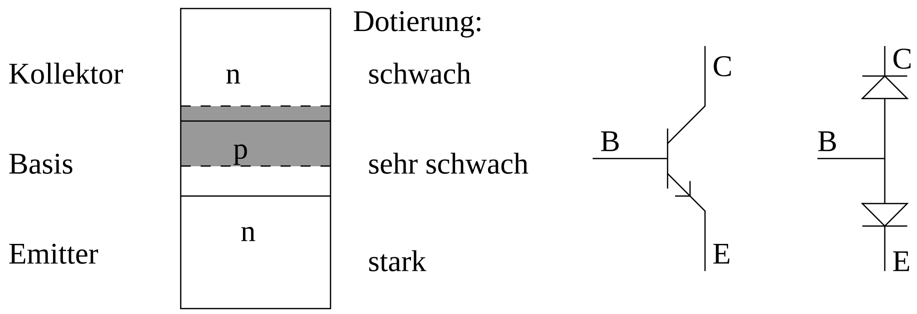
Zenerdioden eignen sich auch für die Verschiebung eines Signales um einen festen Spannungswert (*Levelshifting*), sowie für die Ausgangsspannungsbegrenzung von Verstärkern, wofür man Zenerdioden in den Gegenkopplungszweig schaltet. (Siehe Kapitel 3.6).

```
Weitere Diodenschaltungen:
- Levelshifter Schaltung
- Eingangsueberspannungsschutz
- Cockcraft-Walton oder Booster Schaltung zur HV Erzeugung.
    (fuer HV fuer TV Roehren, PM's usw.)

- Abw"artswandler, Durchflusswandler, Kories Seite 489
    (statt analoge L"angsgeregelte Stromversorgungen)
- Aufwaertswandler, Sperrwandler, Kories, Seite 491.
    (fuer Batteriegeraete, Fotoblitze, usw.)
```

4.11.98

2.4 Bipolare Transistoren

Das Halbleiterelement *Transistor* besteht aus 3 Elektroden, es dient zum Verstärken oder Schalten eines Signales. Speziell versteht man unter *bipolaren Transistoren* solche, die aus 3 unterschiedlich dotierten Halbleiterzonen bestehen, wobei man entsprechend der Dotierungsfolge zwischen npn und pnp Transistoren unterscheidet:



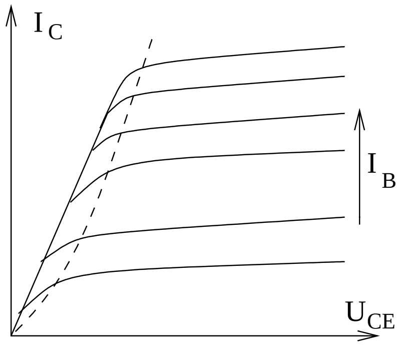
Die beiden pn Übergänge verhalten sich einerseits wie zwei entgegengesetzte Dioden, diese Beschreibung (ganz rechts) gibt aber noch nicht die gesamte Funktionalität wieder.

Normalerweise wird die Basis–Emitterdiode in Vorwärtsrichtung und die Kollektor–Basisdiode in Sperrrichtung betrieben. Dadurch driften Elektronen vom stark dotierten Emitter in die Basisschicht. Da diese sehr schwach dotiert, und ausserdem sehr dünn ist, reicht die Sperrschicht der BC Diode nahe an die CE Übergangschicht heran, sodass viele Elektronen in die BC Sperrschicht driften können. Dort werden sie durch das E-Feld Richtung Kollektor abgesaugt und erzeugen so einen CE Strom. Der zentrale Punkt ist nun, dass dieser Strom durch die Spannung an der BE Diode gesteuert werden kann; diese ist gegeben durch den Basisstrom und die Diodenkennlinie des BE pn Überganges.

Der bipolare Transistor beschreibt man deshalb am zweckmässigsten als Stromverstärker. Das Verhältnis

$$\beta = \frac{\partial I_C}{\partial I_B} \tag{2.14}$$

heisst die *Stromverstärkung* des Transistors. Typische Werte liegen zwischen 10 (Leistungsstransistoren) und 1000 (Kleinsignaltransistoren), je nach Typ. β schwankt stark von Indivivduum zu Individuum und nimmt mit grösserer Kollektor–Basis Sperrspannung U_{CE} zu, da dann die Sperrschicht dicker wird, und somit näher an den BE Übergang herankommt (*Early–Effekt*).



Die sich daraus ergebende Ausgangskennlinie zeigt das nebenstehende Bild. Im normalen Betrieb rechts der gestrichelten Linie ist $U_{CB} > 0$, die Kollector–Basis Diode also gesperrt. Der Kollektorstrom wird vom Basisstrom gesteuert und hängt nur leicht von der CE Spannung ab, der differentielle Ausgangswiderstand ist also klein. Links von der gestrichelten Linie beginnt die CB Diode zu leiten, der Kollektorstrom wird nun unabhängig vom Basistrom, es findet keine Steuerung mehr statt. Man spricht vom *Sättigungsbetrieb.*

Der differentielle Eingangswiderstand r_{BE} ist durch die Diodengleichung der Basis–Emitter Strecke gegeben, Differenzieren von Gleichung 2.7 ergibt

$$r_{BE} = \frac{\partial U_{BE}}{\partial I_B} = \frac{kT}{eI_B} \tag{2.15}$$

Der differentielle Aausgangswiderstand wird (experimentell)

$$r_{CE} = \frac{\partial U_{CE}}{\partial I_C} = \frac{U_\gamma}{I_C} \tag{2.16}$$

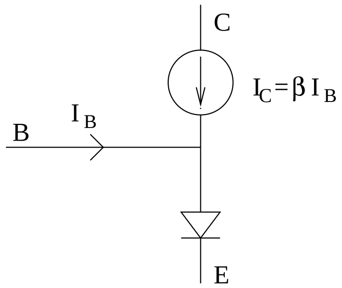
und ist also umgekehrt proportional zum Kollektorstrom. Die Konstante U_γ beschreibt den Early–Effekt und heisst demnach *Earlyspannung.*

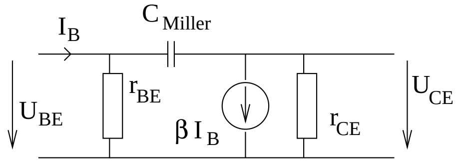
Die Definition der *Steilheit S* (forward transconductance) beschreibt die Steuereigenschaft des Transistors auf eine andere Art, nämlich die Änderung des Kollektorstromes gegenüber der BE Spannung. Man erhält:

$$S = \frac{\partial I_C}{\partial U_{BE}} \approx \frac{\partial I_C}{\partial I_B} \cdot \frac{\partial I_B}{\partial U_{BE}} = \frac{\beta}{r_{BE}} = \frac{eI_C}{kT} \tag{2.17}$$

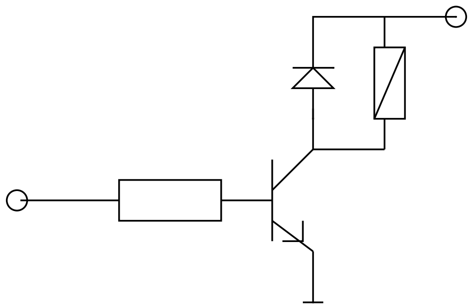
Weiter spricht man von der *Rückwirkung* (backward transconductance) von der CE Spannung auf den Basisstrom. Dieser Effekt ist aber meist klein ausser bei hohen Frequenzen, wo die *Millerkapazität* (effektive Kapazität zwischen Kollektor und Basis, also etwa die BC Grenzschichtkapazität) die Verstärkung reduziert.

Das Grosssignal- (links) und das Kleinsignalersatzschalbild (rechts) illustrieren diese Eigenschaften:





Bipolare Transistoren zeichnen sich durch eine hohe Steilheit, eine realtiv gute Parameterstabilität sowie hohe Geschwindigkeit aus. Nachteilig wirkt sich der relativ hohe Schaltungsaufwand zur Einstellung und Stabilisierung des Arbeitspunktes aus.



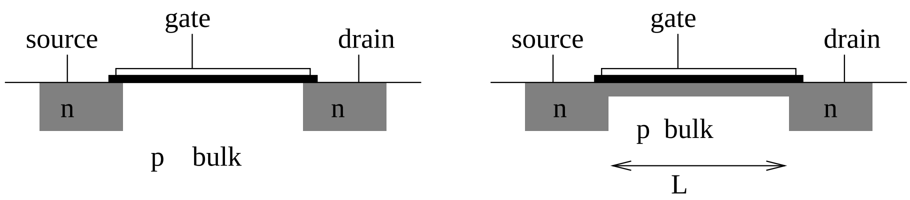
Anwendung: Ansteuerung eines Relais oder einen anderen Stromverbraucher von einem Logiksignal aus. Die Schaltung erlaubt eine separate Speisespannung für das Relais. Die Diode schützt den Transistor vor Überspannungen, die durch die Induktivität entstehen würden.

Wichtige Grenzdaten sind die maximale zulässige Verlustleistung, und die maximalen Sperrspannungen der CE und CB Dioden, wobei erstere oft nur wenige Volt beträgt. Dotierungsstärke und geometrische Grösse des Transistors erlauben die Grenzdaten und die dynmaischen Parameter in einem grossen Bereich zu variieren. Für eine ausführliche Behandlung der verschiedenen Kühlmethoden und ihre Berechnung siehe zum Beispiel in [Nühr98].

2.5 Feldeffekt – Transistoren

Hier handelt es sich um Halbleiter, die im Gegensatz zu bipolaren Transistoren mit einem elektrischen Feld, d.h. leistungslos gesteuert werden, ohne dass in der Steuerelektrode ein nennenswerter Strom fliesst. Man nennt sie *FET*.

Die Steuerung des Stromes im *Kanal* zwischen *Drain* und *Source* erfolgt durch eine Steuerelektrode, das *Gate*.



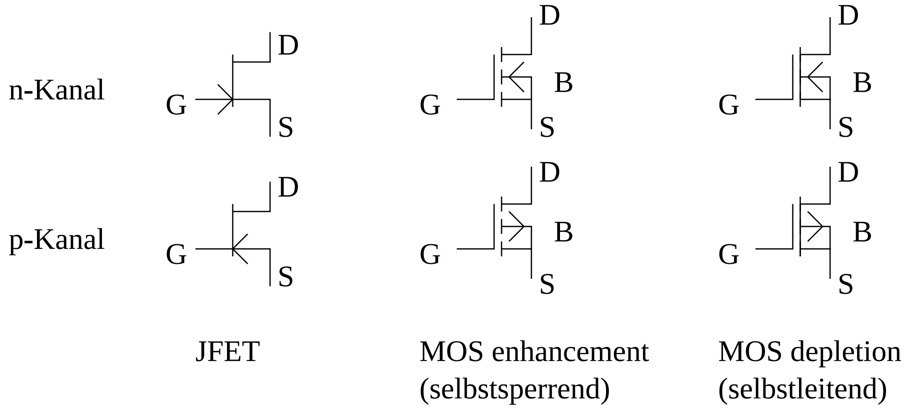
Das Bild links zeigt einen gesperrten n-Kanal FET, der auf einem p dotierten Substrat (bulk) aufgebaut ist. Wird nun an die isolierte Steuerelektrode ("gate") eine positivere Spannung angelegt, werden sich unterhalb des gates Elektronen ansammeln, die vorerst die Löcher des p-Substrates kompensieren, es bildet sich eine Verarmungszone. Wird die Spannung weiter erhöht bildet sich ab der "Threshold voltage" oder "pinch off voltage" V_T ein Überschuss an Elektronen, man erhält unter dem Gate eine zusätzliche n-Schicht, auch *n-Kanal* genannt, die Source und Drain leitend verbindet (Bild rechts). Neben der Dotierungsstärke ist vor allem die geometrische Grösse der gate Elektrode (Länge L, Breite in der nicht gezeichneten Dimension W) ein definierender Faktor für die FET Eigenschaften.

Bei kleinen Drain–Source Strömen lässt sich so der Kanalwiderstand steuern. Bei grösseren Strömen ergibt sich durch den Spannungsabfall zwischen source und drain ein ortsabhängiges elektrisches Feld, das gegen Drain hin immer kleiner wird, der Kanal wird deshalb immer dünner. Damit wird der Strom begrenzt, es ergibt sich ein Sättigungseffekt. Die DS Spannung, bei der die Sättigung beginnt, heisst Kniespannung (U_K). Offensichtlich ist $U_K = U_{GS} - U_T$.

Die Schwellwertspannung V_T hängt neben der Dotierungsstärke auch von der Spannung V_{BS} zwischen bulk und source ab, da diese Spannung die Dicke der Sperrschicht zwischen dem Kanal und dem Bulk und somit die Kanaldicke mitbestimmt.

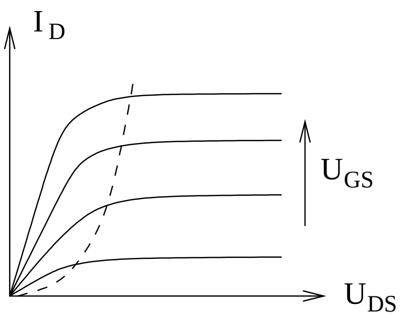
Man unterschiedet *SperrschichtFET* (junction FET, JFET) und *MOSFET*, ("Metal Oxide on Semiconductor Field Effect Transistor", auch MOST). Während bei Sperrschicht FETs eine normale pn Sperrschicht die Isolation zwischen Gate und Kanal darstellt, hat eine MOSFET eine dünne SiO_2 Schicht als Isolator, was einen viel höheren Isolationswiderstand ergibt. MOSFET gibt es selbstleitend (Depletion) und selbstsperrend (Enhancement), je nach dem ob bei Gatespannug null der Kanal leitet oder nicht:

9.11.98



MOSFET sind fast immer selbstsperrend. Selbstleitende MOSFETs erzeugt man durch explizite Dotierung der Kanalzone. Sperrschicht FETs sind immer selbstleitend. Es gibt bei allen drei Sorten p und n Kanal FETs, n-Kanal MOSFETs werden oft kürzer einfach als NMOS bezeichnet (entsprechend PMOS).

Da man in konkreten Schaltungen sowohl NMOS als auch PMOS Transistoren verwenden will, werden oft beide Transistoren auf denselben wafer aufgebracht. Hat man einen p-Substrat und will ein PMOS Transistor machen, implantiert man zuerst eine n dotierte Wanne, die den bulk des PMOS ausmachen wird. NMOS Transistoren können direkt das Substrat als bulk nehmen. Diese Technik wird als "complementary" MOS Technik bezeichnet (*CMOS*). Der grösste Teil aller heute hergestellten Elektronik besteht aus CMOS Technik.



Die nebenstehende Figur zeigt eine typische Ausgangskennlinie. Links der *Kniespannung* (gestrichelten Linie) bestimmt die Gatespannung den Widerstand, rechts davon befindet sich der Sättigungsbereich, wo der Strom fast konstant wird, und der FET also als gesteuerte Stromquelle arbeitet. Bei gegebenen Materialkonstanten (Herstellungsprozess) und Dotierung können die Kennlinien durch Länge L und Breite W des FET beeinflusst werden.

Im ohmschen Bereich gilt ungefähr

$$I_D = k \cdot \frac{W}{L} \cdot ((V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2) \qquad (2.18)$$

(k Materialkonstante). Der Ausgangswiderstand r_{DS} wird also ungefähr

$$r_{DS} := \frac{\partial V_{DS}}{\partial I_D} \approx \frac{1}{k}\frac{L}{W}\frac{1}{V_{GS} - V_T} \qquad (2.19)$$

Das Verhältnis L/W und die Gatespanung bestimmen also den Widerstand. Beachte, dass die Kapazität zwischen Gate und Kanal durch $W \cdot L$ bestimmt ist.

Die Sättigung beginnt sich ab der Kniespannung $V_{DS} = V_K := V_{GS} - V_T$ bemerkbar zu machen. Da ist die Kanalbreite am Drainende gerade auf null gesunken ($V_{GD} := V_{GS} - V_{DS} = V_T$). Die gestrichelte Linie in der Kennlinie ist also durch diese Bedingung definiert. Im Sättigungsbereich gilt:

$$I_D = \frac{k}{2}\frac{W}{L}(V_{GS} - V_T)^2\,(1 + \lambda V_{DS}) \tag{2.20}$$

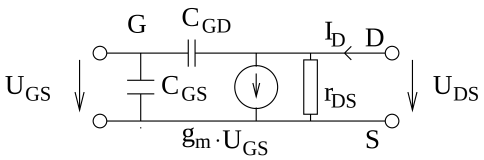
wo λ eine experimentell zu bestimmende kleine Korrekturkonstante ist, die direkt den Ausgangswiderstand bestimmt:

$$r_{DS} := \frac{\partial V_{DS}}{\partial I_D} \approx \frac{1}{\lambda I_D} \tag{2.21}$$

Analog zu der Situation bei den bipolaren Transistoren definiert man die Steilheit S (englisch *Transconductance* g_m) und erhält aus Gleichung 2.20 für den Sättigungsbereich:

$$S := g_m := \frac{\partial I_D}{\partial U_{GS}} = 2\frac{k}{2}\frac{W}{L}(V_{GS} - V_T) = 2\sqrt{\frac{k}{2}\frac{W}{L}I_D} \tag{2.22}$$

Die Transkonduktanz geht also mit der Wurzel des Drainstromes, im Gegensatz zum bipolaren Transistor, bei dem die Steilheit proportional zu I_C war. Im ohm'schen Bereich ist g_m proportional zu V_{DS} wie man aus Gleichung 2.18 ersehen kann.



Das nebenstehende Ersatzschaltbild zeigt die wichtigsten Komponenten eines FET.

Eine vollständigere Beschreibung der Verhältnisse und eine genaue Ableitung der Formeln findet man in [Laker94], Kapitel 1.

Vergleich FET – bipolar

Der wichtigste Unterschied ist, dass der FET eine sehr grossen, vernachlässigbaren Eingangswiderstand hat.

Auch wenn sich die Kennlinien auf den ersten Blick ähnlich sehen, gibt es doch signifikante Unterschiede: Im Sättigungsbereich steigen die Kurven beim FET weniger (grösserer Ausgangswiderstand, also bessere Stromquelle). Dafür lassen sich mit bipolaren Transistoren grössere Verstärkungen erreichen. Die Gegenkopplungskapazität (Millereffekt) ist bei Transistoren allerdings in der Regel höher, sodass sich mit FET höhere maximale Frequenzen erreichen lassen.

Im Bereich kleiner Ausgangsspannungen verhält sich der FET wie ein steuerbarer ohm'scher Widerstand, wohingegen der bipolare Transistor in Sättigung geht, und nicht mehr linear steuerbar ist (nur noch ein–aus).

Da wir es mit Bauelementen zu tun haben, die durch je eine Ein- und Ausgangsspannung und -Strom charakterisiert sind, die alle miteinander zusammenhängen, ist es naheliegend, das Kleinsignalmodell (Linearisierung im Arbeitspunkt) in Matrixschreibweise darzustellen (man spricht auch von Vierplogleichungen). Die sogenannte Y – Matrix (Leitwertmatrix) der bipolaren Transistoren lautet:

$$\begin{pmatrix} \mathrm{d}I_B \\ \mathrm{d}I_C \end{pmatrix} = \begin{pmatrix} \frac{1}{r_{BE}} & 0 \\ S & \frac{1}{r_{CE}} \end{pmatrix} \cdot \begin{pmatrix} \mathrm{d}U_{BE} \\ \mathrm{d}U_{CE} \end{pmatrix} \tag{2.23}$$

Alternativ wird die sogenannte h Darstellung (Hybriddarstellung) verwendet:

$$\begin{pmatrix} \mathrm{d}U_{BE} \\ \mathrm{d}I_C \end{pmatrix} = \begin{pmatrix} r_{BE} & 0 \\ \beta & \frac{1}{r_{CE}} \end{pmatrix} \cdot \begin{pmatrix} \mathrm{d}I_B \\ \mathrm{d}U_{CE} \end{pmatrix} \tag{2.24}$$

(Beachte, dass $\beta = S \cdot r_{BE}$)

Für die FET wird $r_{GS} \to \infty$, deshalb machen nur die folgenden beiden Darstellungen einen Sinn:

$$\mathrm{d}I_D = g_m \cdot \mathrm{d}U_{GS} + \frac{1}{r_{DS}} \cdot \mathrm{d}U_{DS} \tag{2.25}$$

und

$$\mathrm{d}U_{DS} = r_{DS} \cdot \mathrm{d}I_D - \mu \cdot \mathrm{d}U_{GS} \tag{2.26}$$

wobei $\mu = g_m \cdot r_{DS}$ die Spannungsverstärkung bei konstantem Drainstrom bedeutet.

Heute werden bipolare Transistoren nur noch in einzelnen Spezialanwendungen, unter anderem für Stromverstärkungen bei Schaltanwendungen und in Leistungsverstärkern eingesetzt. Der überwiegende Teil der aktiven analogen und digitalen Schaltungen wird gesamthaft auf ein einziges Substrat in sogenannten *integrierten Schaltkreisen* in CMOS Technologie hergestellt.

2.6 Transistoren – Grundschaltungen

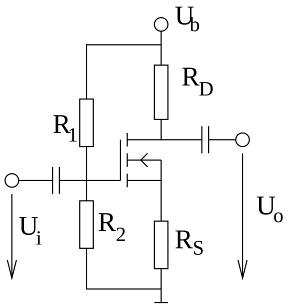
Es gibt entsprechend den 3 Anschlüssen drei Grundschaltungen, wobei der Name jeweils der gemeinsame Referenzpol für Ein- und Ausgangsspannung kennzeichnet: Emitterschaltung (Sourceschaltung), Kollektorschaltung (Drainschaltung), Basisschaltung (Gateschaltung).

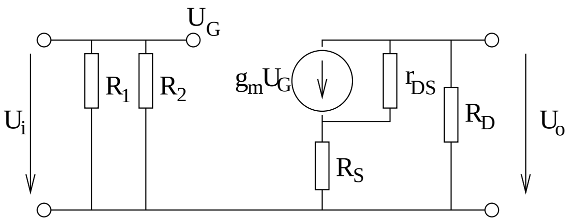
Bei allen Schaltungsarten müssen wir darauf achten, einen geeigneten Arbeitspunkt zu stabilisieren, was einen zusätzlichen Schaltungsaufwand bedeutet. Entsprechend ihrer Bedeutung werden die folgenden Schaltungen jeweils hauptsächlich für MOS Transistoren diskutiert.

2.6.1 Emitter- oder Sourceschaltung

Die *Emitterschaltung* (bei FET *Sourceschaltung*) erlaubt die grösste Verstärkung und entspricht direkt den im letzten Abschnitt diskutierten Funktionsprinzipien der Transistoren.

Da die Wahl eines günstigen Arbeitspunktes eine bestimmte Gleichspannung impliziert, muss das Signal sowohl am Eingang als auch am Ausgang mit Kondensatoren eingekoppelt werden (AC–Kopplung), was zusammen mit den Ein- bzw. Ausgangswiderständen einen effektiven Hochpass bedeutet. Das Ersatzschema gilt für hohe Frequenzen im Vergleich mit der unteren Grenzfrequenz der AC–Kopplung und gleichzeitig für niedere Frequenzen im Vergleich zu den aus der Streukapazität C_{GD} entstehenden oberen Grenzfrequenzen, es enthält deshalb keine Kondensatoren:

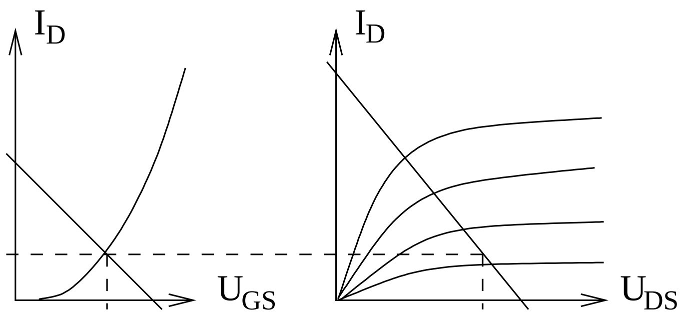




Die beiden Widerstände R_1 und R_2 auf der Eingangsseite legen zusammen mit dem durch den Sourcewiderstand R_S fliessenden Drainstrom den **Arbeitspunkt** fest. Er wird für Verstärkungsanwendungen so gewählt, dass der Transistor im oberen Bereich der Kennlinien (Sättigungsbereich) arbeitet. Es gilt offensichtlich für den Arbeitspunkt:

$$U_{GS} = U_b \frac{R_2}{R_1 + R_2} - I_D \cdot R_S \qquad (2.27)$$

$$U_{DS} = U_b - I_D \left(R_S + R_D\right) \qquad (2.28)$$



Diese Formeln können als gerade in die Kennlinienbilder eingezeichnet werden (links Übertragungskennlinie, rechts Ausgangskennlinie

Bei FETs kann R_1 und R_2 sehr gross gewählt werden, sodass ein hoher Eingangswiderstand entsteht. Falls das Eingangssignal eine genügende DC Stabilität hat, kann die gesamte Eingangsbeschaltung weggelassen werden, was den höchsten Eingangswiderstand erlaubt.

Bei bipolaren Transistoren muss darauf geachtet werden, dass der Querstrom durch diese Widerstände gross gegenüber dem Basisgleichstrom wird. Man wählt R_1 und R_2 ungefähr 10 mal so gross wie der Emitterwiderstand, womit der Basistrom bei einer Stromverstärkung von 100 einen Zehntel des Querstromes ausmacht.

Aus dem Kleinsignalersatzschaltbild können wir nun die Beziehungen für den (differentiellen!) **Eingangswiderstand** ablesen. Das Kleinsignalbild für den bipolaren Fall braucht dabei zusätzlich den Eingangswiderstand r_{BE} vom Eingang zu R_S. Es ergibt sich dann (links bipolar, rechts FET):

$$r_i = \frac{U_i}{I_i} = R_1 \| R_2 \| (r_{BE} + \beta R_S) \qquad r_i = R_1 \| R_2 \tag{2.29}$$

Der **Ausgangswiderstand** $r_o = \mathrm{d}U_o/\mathrm{d}I_o$ berechnet sich in der linearisierten Ersatzschaltung am leichtesten als Verhältnis von Leerlaufspannung durch Kurzschlussstrom, man erhält:

$$r_o = R_C \| (R_E + r_{CE}) \qquad r_o = R_D \| (R_S + r_{DS}) \tag{2.30}$$

Die **Spannungsverstärkung** $A = \mathrm{d}U_o/\mathrm{d}U_i$ wird

$$A = -\frac{g_m R_D}{1 + g_m R_S} \tag{2.31}$$

Lässt man den Sourcewiderstand weg ($R_S = 0$, z.B. durch parallele Kapazität), dann wird die maximale Verstärkung $A_0 = -g_m R_D$ erreicht. Es gilt:

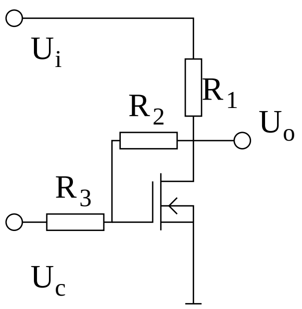
$$\frac{1}{A} = \frac{1}{A_0} - \frac{R_S}{R_D} \tag{2.32}$$

Die Verstärkung wird also durch den Sourcewiderstand reduziert. Wählt man den Sourcewiderstand gross genug, dann dominiert der zweite Term, und die Verstärkung wird gleich $-R_D/R_S$ unabhängig von g_m und damit unabhängig von prozessbedingten Streuungen der Transduktanz. Man spricht von *Stromgegenkopplung.*

Man beachte, dass diese Kleinsignalersatzschaltungen und ihre daraus abgeleiteten Zusammenhänge jeweils zwei wesentliche Annahmen machen: Erstens die Linearisierung um den Arbeitspunkt und zweitens ein bestimmter Frequenzbereich (in unserem Beispiel wurde die Millerkapazität vernachlässigt, die bei hohen Frequenzen die Verstärkung reduziert und eine Phasenverschiebung erzeugt). 11.11.98

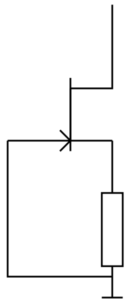
```
Demos:
- Transistor als Schalter (was passiert bei falschem R),
   - L"oschdiode
   - HF Verz"ogerungen
   - Schottki Diode
-Emitterschaltung
   - open loop gain unlinear
   - Grenzfrequenz und Phase wegen Miller
   - Stromgegenkopplung
   - Spannungsgegenkopplung
```

Gesteuerter Spannungsteiler



Der ohm'sche Bereich eines FET kann dafür verwendet werden, einen elektronisch *programmierbaren Spannungsteiler* zu realisieren, der das klassische Potentiometer ersetzen kann. Wie man der Gleichung 2.18 entnehmen kann, hängt der Widerstand U_{DS}/I_D bei grösseren Amplituden noch etwas von U_{DS} ab. Durch Addition der Hälfte der U_{DS} zur Gatespannung ($R_2 = R_3$)wird dieser Effekt korrigiert, man spricht vom *Linearisierungstrick* ([Horo97] p 139, [Kori98] p. 334).

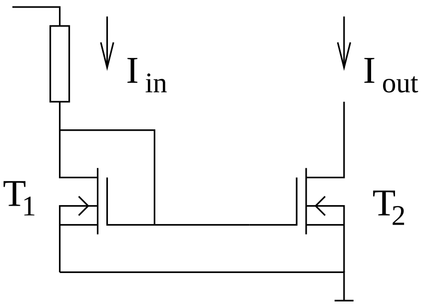
Konstantstromquelle



Stromquellen werden am einfachsten wie nebenstehendes Schaltbild mit selbstleitenden FET's realisiert. Es wird keine Referenzspannung benötigt. Es gilt einerseits $U_{GS} = -R \cdot I$, andererseits liegt der Arbeitspunkt auf der Übertragungskennlinie $I_D(U_{GS})$. Dadurch ist der Strom festgelegt, allerdings gibt es entsprechend der Exemplarstreuung eine relativ grosse Unsicherheit im Wert. Aus dem Ersatzschaltbild liest man folgenden Ausgangswiderstand ab ([Kori98], Seite 331):

$$r = r_{DS} \cdot (1 + g_m R) + R \tag{2.33}$$

Stromspiegel

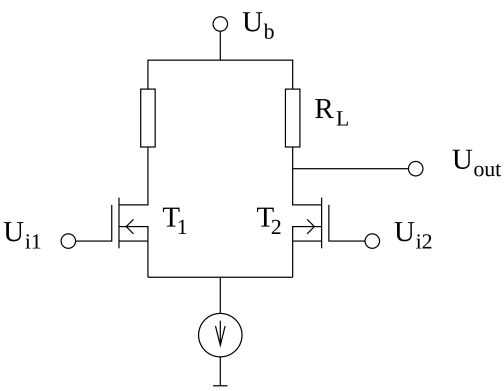


Der rechte FET T_1 sieht einen bestimmten Strom, dieser kann zum Beispiel mit dem eingezeichneten Widerstand eingestellt werden. Da die beiden U_{GS} gleich sind, müssen für zwei identische FET's auch die Ströme in den DS Kanälen gleich sein. Das kann mit grosser Genauigkeit erreicht werden, wenn die beiden FET's unmittelbar nebeneinder auf das gleiche Substrat angebracht werden. (gleiche Prozessparameter, gleiche Betriebstemperatur)

Wählt man die W/L der beiden FET's unterschiedlich, erhält man eine Übersetzung des Stromverhältnisses, es wird ([Laker94], page 378)

$$\frac{I_{out}}{I_{in}} = \frac{(W/L)_2}{(W/L)_1} \tag{2.34}$$

Differenzverstärker



Die Stromquelle an den Source verteilt sich gleichmässig auf die beiden FET's, wenn die Gatespannungen gleich sind. Andernfalls verteilt sie sich im Verhältnis der Differenz der beiden Gatespannungen. An den Lastwiderständen kann also eine Spannung abgegriffen werden, die proportional der Differenz der Eingangsspannung ist ([Kori98], Seite 333). Die *Differenzverstärkung* wird:

$$A_d = \frac{U_o}{U_{i1-U_{i2}}} = -\frac{1}{2} \cdot g_m \cdot R_L \tag{2.35}$$

Weniger Freude hat der Designer an der sogenannten *Gleichtaktverstärkung* A_{CM} ("common mode amplification"). Da die Stromquelle einen endlichen Innenwiderstand r_q hat, wird bei gleichphasiger Bewegung der U_i und somit der gemeinsamen Sourcespannung der Strom sich leicht ändern. Man erhält

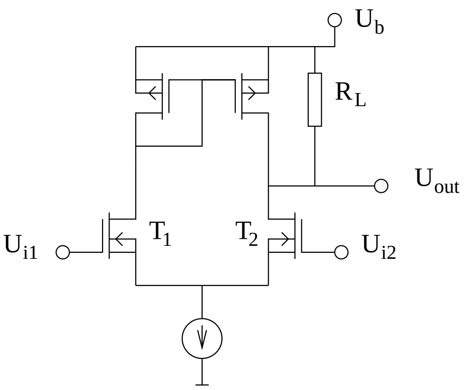
$$A_{CM} = \frac{U_o}{U_{i1}} = -\frac{R_L}{2r_q} \tag{2.36}$$

Das Verhältnis von Differenzverstärkung und Gleichtaktverstärkung heisst *Gleichtaktunterdrückung* oder *common mode rejection ratio* (CMRR).

$$\mathrm{CMRR} = g_m \cdot r_q \tag{2.37}$$

Die Eingangsimpedanz für langsame Signale ist natürlich beliebig hoch. Für schnelle Signale wirkt die Gate-Channel – Kapazität C_{GC} als Last und bestimmt damit den Eingangswiderstand und reduziert aber auch die Verstärkung durch Gegenkopplung (*Millerkapazität*).

Will man ein Signal im Vergleich zur Masse verstärken, legt man zweckmässigerweise das Gate von T_2 auf Masse. Diese Massnahme hat noch einen interessanten Nebeneffekt: Verzichtet man nämlich gleichzeitig auf den Lastwiderstand von T_1 wird die obere Grenzfrequenz drastisch verbessert. Das liegt daran, dass der Drain von T_1 nun festliegt, und es demnach keine gegenphasige Einkopplung auf das Gate mehr geben kann. Da das Gate von T_2 auf Masse liegt, hat dessen HF–Gegenkopplung ebenfalls keine Wirkung! Es handelt sich eigentlich bei T_1 um eine Drain–, bei T_2 um eine Gateschaltung (siehe unten), die gesamte Schaltung wird vergleichbar mit einer sogenannten *Cascode*, siehe unten.

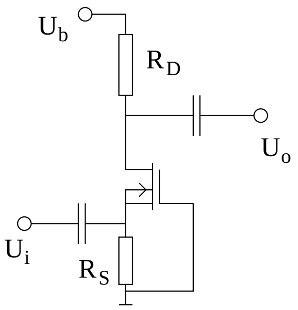


Eine Verbesserung der Gesamtverstärkung bekommt man durch eine *aktive Last* in Form eines Stromspiegels ([Laker94], p. 363). Dabei wird die definierende Seite eines PMOS–Stromspiegels als Last von T_1, die Ausgangsseite als (Teil-)Last von T_2 geschaltet. Eine negative Änderung der Eingangsspannung bewirkt nun gleichzeitig eine Erhöhung des Drainstromes von T_2 und eine Erniedrigung des Ausgangsstromes der Stromquelle auf der Seite von T_2.

Wählt man den Lastwiderstand, der nun die Summe der beiden Stromänderungen aufnehmen muss, gross genug, kann man sehr hohe Verstärkungen von 5000 und mehr erreichen.

Eine wichtige Anwendung des Differenzverstäkers ist der *Komparator*, bei dem zwei Eingangsspannung verglichen werden, und ein Signal “0” oder “1” ausgegeben wird, ja nach dem welcher der beiden Eingänge höher ist als der andere (Soll-Istwert-Vergleich, Schwellwerte für Steuerungen und Alarmsysteme, usw.). Es handelt sich also um eine 1 Bit Digitalisierung. Komparatoren sollen eine hohe Verstärkung und eine schnelle Anstiegzeit haben.

2.6.2 Basis- und Gateschaltungen



Diese Schaltung hat eine Stromverstärkung von etwa 1 und einen niedrigen Eingangswiderstand, scheint also auf den ersten Blick nicht sehr nützlich. Die Hauptanwendung besteht jedoch in der Entkopplung von Ein- und Ausgang und somit dem Unschädlichmachen der *Millerkapazität*. Jeglicher HF Strom durch dies fliesst nach Masse ab, und kann somit keine Gegenkopplung bewirken.

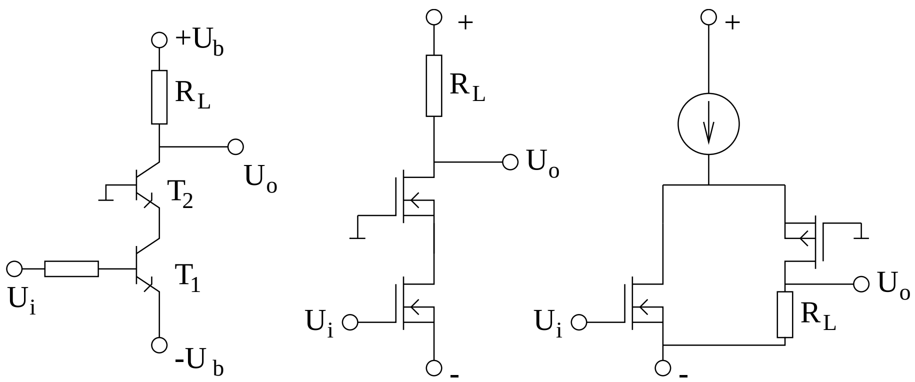
Die Analyse des Kleinsignalersatzschaltbildes liefert

$$r_i = \frac{1}{g_m} \| R_S \qquad A = \frac{U_o}{U_i} = g_m \cdot R_D \tag{2.38}$$

Es besteht also trotz kleinem Eingangswiderstand eine Spannungsverstärkung, die man sich in der HF–Technik zu Nutze macht ([Kori98], p 329).

Kaskodenschaltungen

Die folgenden Skizzen zeigen die klassische Anwendung der Basisschaltung, die Kaskode. Links eine bipolare Kaskode, in der Mitte eine solche aus FET. 16.11.98



Das Signal wird vorerst mit einem Transistor T_1 in Emitterschaltung verstärkt. Da der Kollektor den kleinen Eingangswiderstand des T_2 in Basisschaltung sieht, ändert sich die Spannung am Kollektor praktisch nicht, die Gegenkopplung durch die Millerkapazität ist deshalb vernachlässigbar. Der von T_1 verstärkte Strom wird schliesslich mit Hilfe von T_2 in eine Spannung am Lastwiderstand R_L umgewandelt. Die Gesamtspannungsverstärkung beträgt:

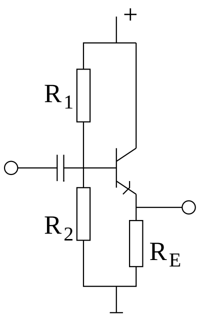
$$A = -g_m \cdot R_L \tag{2.39}$$

Die rechte Skizze zeigt die *gefaltete Kaskode*. Durch die Stromquelle wird die Änderung des Drainstromes von T_1 nun auf T_2 umgelegt. (Bei letzterem handelt es sich in der gefalteten Kaskode um einen Transistor der umgekehrten Kanalpolarität als T_1.) Der Vorteil liegt vorallem darin, dass der HF–Signalstrom nicht zwischen den Betriebsspannungen fliessen muss.

Neben der Verstärkung auch sehr hohen Frequenzen findet die Kaskode noch überall dort Anwendung, wo eine schlecht definierte Rückkopplungskapazität unerwünscht ist (z.B. bei Integratoren).

(siehe auch [Horo97] page 103, [Laker94] page 344)

2.6.3 Kollektor- und Drainschaltungen



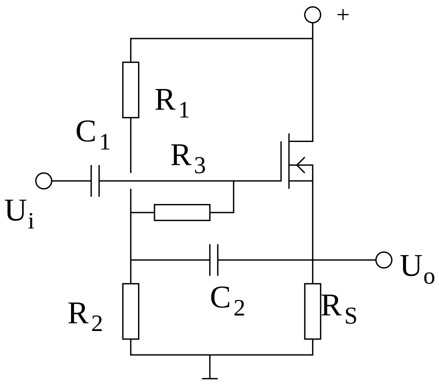
Die Kollektorschaltung wird auch *Emitterfolger* (*Sourcefolger*) genannt, da der Emitter jeweils etwa 0.6 V tiefer als die Basis liegt, und demnach derem Signal mit einem um 0.6 V tieferen Wert folgt. Die Spannungsverstärkung ist hier zwar 1, jedoch gibt es einen grossen Eingangs- und kleinen Ausgangswiderstand, so dass der Emitterfolger als Impendanzwandler eingesetzt wird. Eine typische Anwendung ist ein 50 Ω Kabeltreiber.

Der Ausgangswiderstand eines Emitterfolgers (links) bzw. Sourcefolgers (rechts wird aus der Kleinsignalanalyse bestimmt:

$$r_a \approx \frac{r_{BE}}{\beta} \qquad r_a = \frac{R_S}{1 + g_m R_S} \tag{2.40}$$

([Kori98], p. 303 und 328).

Der Ausgangswiderstand eines Emitterfolgers ist also i.A. kleiner als der Widerstand R_E. Will man bei einem Kabeltreiber für einen richtigen Abschluss sorgen, muss man einen entsprechenden Abschlusswiderstand in Serie zwischen Kabel und Emitter schalten. In Praxis ist die Wahl des Widerstandes allerdings nicht so einfach, da r_a von g_m (bzw. R_{BE} und β) abhängt. Man nimmt deshalb besser Operationsverstärker als Kabeltreiber, falls es auf korrekten Abschluss ankommt.



Die Widerstände R_1 und R_2 legen den Arbeitspunkt fest. Bei FET's sind sie natürlich unerwünscht, da sie den Eingangswiderstand verkleinern. Die sogenannte *Bootstrapschaltung* verschafft Abhilfe, indem das Wechselspannungssignal mit Hilfe von C_2 von der Source auf den Spannungsteiler eingekoppelt wird. Der Signalstrom im Widerstand R_3 verschwindet wegen der Spannugsverstärkung von 1, der Eingangswiderstand der skizzierten Schaltung beträgt etwa 100 MΩ .

Sourcefollower hat unvorhersehbarer Offset (V_{GS} vs I_D ist ziemlich prozess variabel): Zero Voltage follower , [Horo97], p. 135

Leistungsverstärker

[Hin96], Seite 75/78. Emitterfolger im A–Betrieb, Pushpull im B–Betrieb. Saubere Übergabeschaltung Seite 78.

Konstantspannungsquelle

[Horo97] einfach S. 69, reguliert S. 105

Darlingtonschaltung

[Kori98] p. 290

FET analog Schalter

[Horo97] p 143

typisch 25 bis 100 Ω , beeinflusst Schaltgeschwindigkeit.

Achtung, Analogsignal nicht zu nahe an Steuerspannung kommen.

Schaltspitzen wegen kapazitiver Kopplung der Steuerspannung auf die Analogsignale.

Anwendungen: schaltbare Verstärkungsfaktoren; sample and hold; DAC. 18.11.98

Kapitel 3

Signale und Systeme

3.1 Signale

Unter *Signale* versteht man skalare Funktionen der Zeit $f(t)$, die als Strom oder Spannung, aber auch als Licht im Wellenleiter oder als beliebige andere physikalische Grösse vorkommen können. Es gibt periodische und nichtperiodische Signale, deterministische und zufällige Signale. Kausale Signale sind solche, die für alle Zeiten $t < 0$ verschwinden.

Unter Energie eine Signales wird in der Nachrichtentechnik das quadratische Integral

$$E = \int\limits_{-\infty}^{+\infty} |f(t)|^2 \mathrm{d}t \tag{3.1}$$

verstanden, Signale mit nicht unendlicher Energie heissen Energiesignale.

Unter Leistung eine Signales wird das auf die Zeit normierte quadratische Integral

$$P = \lim_{T\to\infty} \frac{1}{2T} \int\limits_{-T}^{+T} |f(t)|^2 \mathrm{d}t \tag{3.2}$$

verstanden, Signale mit endlicher Leistung heissen Leistungssignale.

Energiesignale haben $P = 0$, Leistungssignale $E = \infty$. Energiesignale sind in der Zeit "endlich lang" (z.B. $e^{-t/\tau}$ für $t \geq 0, t = 0$ sonst), Leistungssignale sind "unendlich lang" (z.B. harmonische Funktion). Alle periodischen Signale sind Leistungssignale, haben also $E = \infty$.

Für Untersuchungen von **periodischen** Signalen werden diese in Fourierreihen zerlegt, und die harmonischen Komponenten in der Elektronik oft einzeln betrachtet, man spricht von Frequenzraumanalyse. Da bei jeder Komponente Amplitude und Phase relevant sind, werden sie meist als komplexe Zahlen geschrieben, wie das im ersten Kapitel behandelt wurde.

Beispiele für periodische Signale (n = Fourierreihenindex, $\omega = 2\pi/T$ = Kreisfrequenz):

- positive Rechteckimpulse: Periode T, Pulsdauer τ, *Tastverhältnis* $v := \tau/T$. Fourierkomponenten:
$$f(t) = v + \frac{2}{\pi} \sum \frac{\sin(n\pi v)}{n} \cos(n\omega t) \tag{3.3}$$
Für $v = 0.5$ treten nur die ungeraden Komponenten auf.

- symmetrische Trapezschwingung mit Periode T, Anstiegszeit = Abfallzeit = τ, $a = \pi\tau/T$. Fourierkomponenten:
$$f(t) = \frac{4}{a\pi} \sum_{n=1,3,5...} \frac{\sin(na)}{n^2} \cos(n\omega t) \tag{3.4}$$

Man sieht also insbesondere, dass die Übertragung einer Trapezschwingung weniger hohe Bandbreite braucht, als eine solche für eine Rechteckschwingung.

Für Anwendungen mit einzelnen, **nichtperiodischen** Signalen diskutiert man die Antwortsignale eines Systems im Zeitraum. Dafür verwendet man die Technik der Laplacetransformation.

Beispiele für einzelne Signale im Zeitraum:

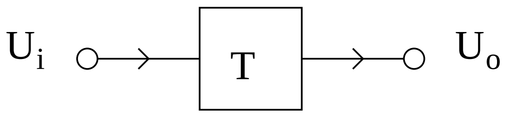
- Unter *Impuls* versteht man ein endliches Signal das ausserhalb eines eingeschränkten Zeitbereichs verschwindet. z.B. einelner Rechteckimpuls, Dreieckimpuls, Gaussimpuls. Der Impuls ist ein Energiesignal.

- Die *Stossfunktion* oder der *Stossimpuls* (Dirac'sche Deltafunktion) ist definiert durch
$$\int_{-\infty}^{+\infty} f(t)\, \delta(t - t_0)\, \mathrm{d}t = f(t_0) \tag{3.5}$$
Sie hat eine endliche Fläche von 1, es handelt sich also um ein Energiesignal.

- Die *Sprungfunktion* $s(t)$ (bis zu $t = 0$ ist $s(t) = 0$, für $t \geq 0$ ist $s(t) = 1$). $s(t)$ ist ein Leistungssignal, $E = \infty$.

Die "Ableitung" der Sprungfunktion liefert den Dirac–Impuls.

3.2 Systeme



Ein *System* ist eine Abbildung T, die ein Signal (das Eingangssignal U_i) in ein anderes (das Ausgangssignal U_o, die Systemantwort) verwandelt:

$$U_0 = T(U_i) \tag{3.6}$$

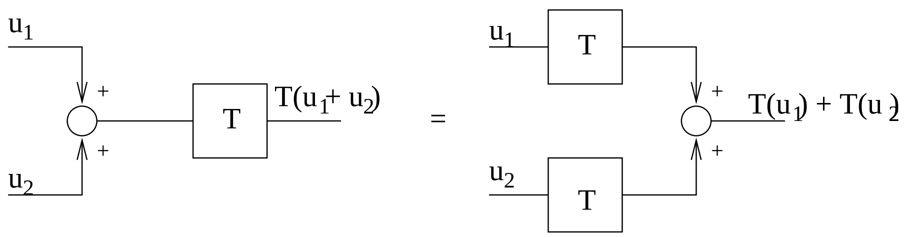
Kausale Systeme sind solche, die keine Systemantwort geben, bevor das Signal nicht beginnt. Anders ausgedrückt: kausale Systeme wandeln kausale Signale wieder in solche um.

3.3 LTI Systeme

Lineare Systeme sind solche für die gilt

$$T(u_1 + \alpha u_2) = T(u_1) + \alpha T(u_2) \tag{3.7}$$

also die gleiche Definition wie in der Linearen Algebra. In der Systemtheorie bezeichnet man dies auch als Superpositionsprinzip oder Überlagerungssatz, es lässt sich grafisch darstellen:



Zeitinvariante Systeme reagieren auf das Eingangssignal unabhängig vom Zeitpunkt seines Eintreffens. Es gilt

$$\text{falls } y(t) = T(x(t)) \quad \text{dann } y(t - t_0) = T(x(t - t_0)) \tag{3.8}$$

Lineare und zeitinvariante Systeme heissen LTI Systeme (linear and timeinvariant). Realisierbare LTI Systeme sind zudem immer kausal.

Systeme, die aus Widerständen, Induktivitäten, Kapazitäten und linearisierten gesteuerten Quellen (z.B. Transistoren im Kleinsignalersatzschaltbild) bestehen, sind LTI Systeme, vorausgesetzt, die Übertragungsfunktion bleibt endlich ("es schwingt nicht").

3.3.1 Betrachtungen im Zeitraum

Die *Sprungantwort* $h(t)$ ist das Ausgangssignal eines Systems, das mit der Sprungfunktion angeregt worden ist. Beispiel: Die Sprungantwort eines Tiefpasses ist $h(t) = 1 - e^{-t/RC}$.

Die *Impulsantwort* oder *Stossantwort* $g(t)$ ist das Ausgangssignal eines Systems, das mit einem Diracstoss angeregt worden ist. Es gilt also $g(t) = T(\delta(t))$, $g(t)$ wird auch *Gewichtsfunktion* genannt. Beispiel: Die Impulsantwort des RC Tiefpasses ist $e^{-t/RC}$ für $t > 0$.

In linearen Systemen bleibt Ableitung und Integral erhalten. Die Ableitung der Sprungantwort ist also die Impulsantwort.

Die Antwort auf ein *beliebiges Eingangssignal* $x(t)$ ergibt sich durch folgende anschauliche Überlegung ([Hin96], Seite 2): Man zerlege das Signal in viele sich folgende gewichtet einzelne Impulse $x(t) = \int x(\tau)\delta(\tau - t)\mathrm{d}\tau$. Die $x(\tau)$ sind also die Gewichte (Amplituden) der Deltafunktion, die zur Zeit τ aktiv ist. Wegen der Linearität vertauschen das Integral und die Abbildung T des Systems, es gilt:

$$T\left(\int x(\tau)\delta(\tau - t)\mathrm{d}\tau\right) = \int x(\tau)T(\delta(\tau - t))\mathrm{d}\tau \tag{3.9}$$

Wegen der Zeitinvarianz gilt

$$g(\tau - t) = T(\delta(\tau - t)) \quad \Leftrightarrow \quad g(t) = T(\delta(t)) \tag{3.10}$$

Wir können also behaupten, dass ein LTI–System T mit der Impulsantwort $g(t)$ auf ein beliebiges Eingangssignal $x(t)$ mit einem Ausgangssignal $y(t)$ reagiert, das durch das zweiseitige *Faltungsintegral* gegeben ist.

$$y(t) = \int\limits_{-\infty}^{+\infty} x(\tau) \cdot g(t - \tau)\, \mathrm{d}\tau =: x(t) \star g(t) \tag{3.11}$$

womit auch der Faltungsoperator $\star$ definiert ist. Die Faltung ist eine lineare Operation, es gelten das Kommutativ-, das Assoziativ und das Distributivgesetz (Faltung und Addition vertauschen). Siehe [Kori98], Seite 240 ff.

Die Fouriertransformierte der Stossantwort ist gerade die *Übertragungsfunktion* $G(\omega)$: Berechnet man nämlich die Antwort eines Systems auf die harmonische Anregung $x(t) = e^{i\omega t}$ erhält man:

$$y(t) = \int\limits_{-\infty}^{+\infty} e^{i\omega\tau} g(t - \tau)\, \mathrm{d}\tau = \int\limits_{-\infty}^{+\infty} e^{i\omega(t-\tau)} g(\tau)\, \mathrm{d}\tau = e^{-i\omega t} \int\limits_{-\infty}^{+\infty} e^{-i\omega\tau} g(\tau)\, \mathrm{d}\tau = x(t) \cdot G(\omega) \tag{3.12}$$

Wir hatten im ersten Kapitel gesehen, wie man mit Hilfe des komplexen ohm'schen Gesetzes die Übertragunsfunktion berechnen kann. Hier sehen wir also, dass entweder die Stossantwort, die Sprungantwort oder die Übertragungsfunktion ausreicht, um das System vollständig zu beschreiben.

3.3.2 Betrachtungen im Frequenzraum

LTI Systeme reagieren auf harmonische Eingangssignale mit harmonischen Eingangssignalen gleicher Frequenz. Es reicht also die Eingangssignale in ihre Fourierreihe zu zerlegen, und die Systemantwort der einzelnen Frequenzkomponenten zu untersuchen.

Etwas formaler: Die Faltung zweier Funktionen geht durch die Fouiertransformation in eine Multiplikation über ([Bron98], Seite 727):

$$y(t) = g(t) \star x(t) \quad \Leftrightarrow \quad Y(\omega) = G(\omega) \cdot X(\omega) \tag{3.13}$$

wobei Grossbuchstaben die Fouriertransformierten bezeichnen. Man kann also die Reaktion eines Systems im Frequenzraum dadurch berechnen, indem man das Eingangssignal fouriertransformiert, dann mit der Übertragungsfunktion multipliziert. Das Ausgangssignal erhält man dann durch die inverse Fouriertransformation.

Setzt man als speziellen Fall $x(t) = \delta(t)$, berechnet sich das Faltungsintegral gerade zu $g(t)$, die Stossantwort. Die Fouriertransformation der Deltafunktion ist eine Konstante, es bleibt also in der Tat die Übertragunsfunktion als Fouriertransformation der Stossantwort.

3.4 Analyse im Zeitraum: Laplace – Transformationen

Muss man das Verhalten eines Systems im Zeitraum verstehen und optimieren, dann eignet sich dazu Laplacetransformationen:

Das Faltungsintegral, das ein System beschreibt, kann nämlich nur dann mit Hilfe der Fouriertransformation in eine Multiplikation übergeführt werden, wenn die Funktionen gewisse Bedingungen erfüllen. So müssen zum Beispiel sowohl die Funktionen als auch ihr Betrag integrierbar sein ([Bron98], Seite 722). Da dies gerade bei interessanten Signalen im Zeitbereich oft nicht erfüllt ist (zum Beispiel Sprungfunktion!), verwendet man meist eine andere Integraltransformation, nämlich die *Lapalacetransformation*. Sie ist definiert als einseitige Integraltransformation ([Bron98], Seite 708):

$$L(f(t)) := \int_0^\infty e^{-st}\, f(t)\, \mathrm{d}t = F(s) \qquad s = \sigma + i\omega \tag{3.14}$$

s ist die komplexe Frequenzvariable, sie geht für $\sigma \to 0$ in die normale Frequenz $i\omega$ über. Dieser gegenüber der Fouriertransformation (zweiseitige Integraltransformation)

$$\mathcal{F}(f(t)) := \int_{-\infty}^{+\infty} e^{-i\omega t}\, f(t)\, \mathrm{d}t = F(\omega) \tag{3.15}$$

eingeführte Dämpfungsterm σ erweitert die Klasse der transformierbaren Funktionen erheblich. So ist zum Beispiel die Laplacetransformierte der Sprungfunktion gerade $1/s$.

Genauso wie für die Fouriertransformation gilt auch für die Laplacetransformation, dass sich eine Faltung im Originalbereich in eine Multiplikation im Bildbereich verwandelt. Man kann also wie oben für die Fouriertransformation beschrieben, die Antwort eines Systems auf ein Eingangssignal $x(t)$ leicht dadurch berechnen, dass man $x(t)$ laplacetransformiert, dann mit der Übertragungsfunktion $G(s)$ (ω ist durch s ersetzt worden) multipliziert und schliesslich in den Originalbereich zurücktransformiert.

Für diese Operationen stehen Laplacekorrespondenztabellen zur Verfügung, siehe zum Beispiel [Bron98], Seite 1063 ff. Insbesondere ist die Transformierte der Sprungfunktion $1/s$, die der δ Funktion gerade 1. $e^{-\alpha t}$ transformiert sich in $1/(s+\alpha)$. Die Transformierte der Ableitung $y'(t)$ wird $sF(s) - y(0)$. 23.11.98

3.4.1 Pole in der Übertragungsfunktion

Das Verhalten eines Systems lässt sich auch als Differentialgleichung in den Ein- und Ausgangssignalen $x(t)$ und $y(t)$ schreiben, in der Form

$$a_n x^{(n)}(t) + a_{n-1} x^{(n-1)}(t) + \ldots + a_0 x(t) = b_n y^{(n)}(t) + b_{n-1} y^{(n-1)}(t) + \ldots + b_0 y(t) \tag{3.16}$$

Die Laplacetransformation von gewöhnlichen Differentialgleichungen n-ter Ordnung führt auf Potenzreihen ([Bron98], Seite 719). Die Übertragungsfunktion lässt sich deshalb als Quotient zweier Potenzreichen schreiben:

$$G(s) = \frac{Y(s)}{X(s)} = \frac{a_n s^n + a_{n-1} s^{n-1} + \ldots + a_1 s + a_0}{b_m s^m + b_{m-1} s^{m-1} + \ldots + b_1 s + b_0} = k \frac{(s-\alpha_n)(s-\alpha_{n-1})\ldots(s-\alpha_0)}{(s-\beta_n)(s-\beta_{n-1})\ldots(s-\beta_0)} \tag{3.17}$$

Die n Nullstellen α_i und vorallem die m Pole β_i haben eine konkrete Bedeutung für die Beurteilung der Stabilität eines Systems, sowie dessen Einschwingvorgängen. Nach dem Dämpfungssatz der Laplacetransformation ist nämlich $L(f(t)) = F(s) \Leftrightarrow L(e^{s_0 t} f(t)) = F(s - s_0)$ (Verschiebung nach links, [Bron98], Seite 709), also gilt:

1. Jeder Pol der Übertragunsfunktion an der Stelle $s = s_0$ führt zu einem Einschwingvorgang (Stossantwort) der Form $e^{s_0 t}$.
2. Ein System ist genau dann stabil (schwingt nicht, divergiert nicht), wenn alle Pole in der linken Hälfte ($\sigma < 0$) der s–Ebene liegen.
3. Zu jedem nicht reellen Pol $s_0 = a + ib$ existiert auch der konjugiert komplexe Pol $s_0^\star = a - ib$. (weil die b_i reell sind)

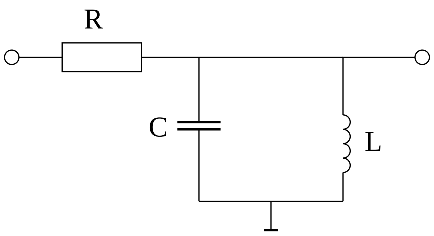
Liegt ein Pol auf der negativen reellen Achse, ergibt sich also ein exponentieller Abfall (kritische Dämpfung). Liegt er an der Stelle $-a \pm ib$ erhält man eine Stossantwort der Form $e^{-at} \sin bt$.

Um das Verhalten eines Systems im Zeitraum zu erforschen, reicht es also die Lage der Pole der Übertragungsfunktion $G(s)$ zu diskutieren. Man erhält daraus direkt die Stossantwort, die wiederum das Systemverhalten vollständig beschreibt.

Beispiel **Tiefpass erster Ordnung** ($i\omega \to s$, $\omega_0 = 1/RC$):

$$G(s) = \frac{1/sC}{R + 1/sC} = \frac{\omega_0}{\omega_0 + s} \tag{3.18}$$

Die Übertragungsfunktion hat also einen einzigen Pol bei $-w_0$ auf der negativen s-Achse. Die Stossantwort lautet $\omega_0\, e^{-\omega_0 t}$.

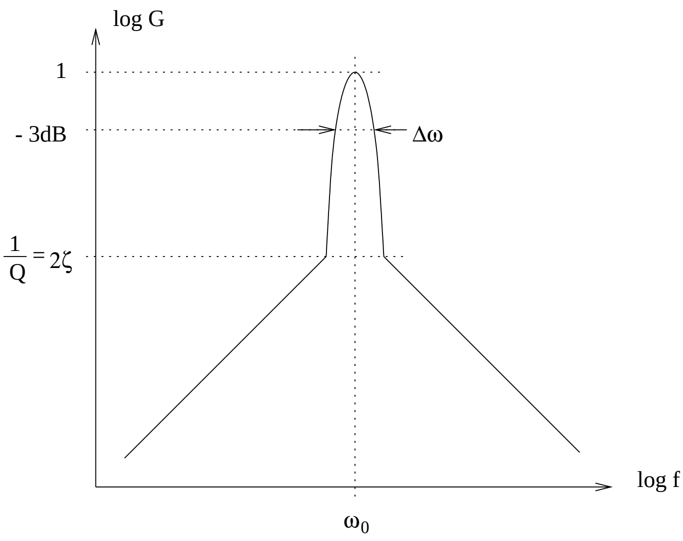


Beispiel RCL Bandpassfilter: Dieses besteht aus einem Tiefpass zweiter Ordnung. Die Übertragunsfunktion berechnet sich als Spannungsteiler mit komplexen Widerständen zu

$$G(s) = \frac{s \cdot L/R}{1 + s \cdot L/R + s^2 LC} \tag{3.19}$$

Wir definieren die Resonanzfequenz $\omega_0 = 1/\sqrt{LC}$ und den Dämpfungsfaktor $\zeta = \omega_0 \cdot L/2R$, damit erhalten wir:

$$G(s) = \frac{2\zeta\omega_0 s}{w_0^2 + 2\zeta\omega_0 s + s^2} \tag{3.20}$$



Für die Diskussion im **Frequenzraum** setzen wir $s = i\omega$ und beobachten, dass für $\omega = \omega_0$ die Übertragunsfunktion gerade 1 wird. Weit oberhalb und unterhalb fällt $|G(\omega)|$ mit $1/\omega$ ab es ergibt sich eine Phasenverschiebung von $\pm 90°$. Um die Resonanzfrequenz herum ergibt sich ein schmaler Peak, mit einer 3 dB Breite von $\Delta\omega := \zeta\omega_0$

Man definiert auch die *Güte* des Bandpasses (analog Schwingkreisgüte):

$$Q := \frac{\omega_0}{2\Delta\omega} = \frac{1}{2\zeta} \tag{3.21}$$

Die Güte ist also eine grosse Zahl, wenn wir einen schmalbandigen Filter bauen. ζ ist gerade die relative halbe -3 dB Breite von $|G(\omega)|$.

Für die Diskussion im Zeitraum müssen wir die Pole in der Übertragungsfunktion diskutieren, d.h. die Lösungen der Gleichung

$$w_0^2 + 2\zeta\omega_0 s + s^2 = 0 \tag{3.22}$$

Diese quadratische Gleichung hat zwei Wurzeln

$$s_{1/2} = \omega_0\,(-\zeta \ \pm \sqrt{\zeta^2 - 1}) \tag{3.23}$$

Ist $\zeta > 1$ dann liegen beide Wurzeln auf der negativen reellen Achse, die Stossantwort enthält $e^{s_1 t} + e^{s_2 t}$. Für $\zeta < 1$ liegen die beiden konjugiert komplexen Lösungen auf einem Kreis mit Radius ω_0. Die Stossantwort enthält dadurch also einen Einschwingterm $e^{s_1 t} + e^{s_2 t} = 2\ e^{-\zeta\omega_0 t}\ \cos(\omega_0\sqrt{1-\zeta^2}\,t)$. Für $\zeta = 1$ wandern die beiden Lösungen zusammen zu $s_0 = -\zeta\omega_0$, wir erhalten einen kritisch gedämpften Einschwingvorgang.

Mit Hilfe der Laplacetransformation können wir die Systemantwort $y(t)$ der Sprunfunktion direkt und vollständig bestimmen, indem wir das Produkt der Übertragungsfunktion $G(s)$ mit der Laplacetransformierten der Sprungfunktion $1/s$ multiplizieren, und das Resultat zurück in den Zeitraum transformieren. Aus der Laplace–Korrespondenztabelle [Bron98], Seite 1063, Formel 7 entnehmen wir:

$$F(s) = \frac{1}{s^2 + 2\alpha s + \beta^2} \quad \Leftrightarrow \quad f(t) = \frac{1}{\sqrt{\beta^2 - \alpha^2}}\ e^{-\alpha t}\ \sin\sqrt{\beta^2 - \alpha^2}\ t \tag{3.24}$$

Unsere entsprechende Werte eingesetzt, ergibt die Sprungantwort für unseren Bandpass von

$$y(t) = \frac{2\zeta}{\sqrt{1-\zeta^2}}\, e^{-\zeta w_0 t}\ \sin \sqrt{1-\zeta^2}\, \omega_0 t \tag{3.25}$$

Dieser Fall kommt vor allem in der Kern- und Teilchenphysik relativ häufig vor: Man will ein Ereignis messen, das in einem Sensor eine Ladungsmenge proportional seiner Energie deponiert hat. Das entspricht einer Sprungfunktion, und wir wollen die Höhe des Sprunges vermessen. Ein derartiger Bandpass macht nun aus jedem Sprung ein schöner Puls, der von der Vorgeschichte relativ unabhängig ist. Wählt man ζ so, dass das Signal möglichst schnell wieder auf null ist, aber fast nicht mehr überschwingt, spricht man auch von einem *Semigauss – Filter* oder einem *Shaper*. Das ist etwa bei $\zeta = 0.8$ der Fall.

Solche Filter können mit Hilfe von Operationsverstärkern für praktisch beliebige Übertragungsfunktionen realisiert werden, nur unter Verwendung von Widerständen und Kapazitäten. Man spricht dann von aktiven Filtern (siehe Kapitel 4.2.5).

3.5 nichtlineare Systeme

Nichtlineare Systeme sind solche, deren Ausgangsgrösse nicht linear von der Eingangsgrösse abhängt. In der Praxis haben alle Systeme nicht lineare Komponenten, z.B. ist die Linearisierung im Arbeitspunkt von Halbleiterelementen nur eine Näherung. Alle praktischen Systeme haben ausserdem Aussteuergrenzen, d.h. Ausgangssignale können nur einen bestimmten Maximalwert erreichen, meist etwas weniger als die Spannungsversorgungswerte.

Nichtlineare Übertragungssysteme werden mit ihrer Kennlinengleichung beschrieben, häufig als Potenzreihe angenähert. In letzterem Fall nennt man die Ordnung des Systems die höchste vorkommende Potenz.

In nichtlinearen Systemen n-ter Ordnung tauchen für harmonische Eingangssignale einer Frequenz ω auch Ausgangssignale der Frequenzen $n \cdot \omega$ auf. (Beispiel $n = 2$, $x(t) = \cos \omega t$, $y(t)$ enthält dann einen Term $\cos 2\omega t$.)

Um die Abweichung eine Systems von der Linearität zu beschreiben, bedient man sich des *Klirrfaktors* eine Systems k. Er definiert durch das Verhältnis der Effektivwerte aller Oberschwingungen und diejenigen des Gesamtsignales, wenn man als Eingangssignal ein harmonisches Signal einer einzigen Frequenz verwendet (A_n Amplitude der

n-ten Oberschwingung):

$$k = \frac{\sqrt{\sum\limits_{n=2}^{\infty} A_n^2}}{\sqrt{\sum\limits_{n=1}^{\infty} A_n^2}} \tag{3.26}$$

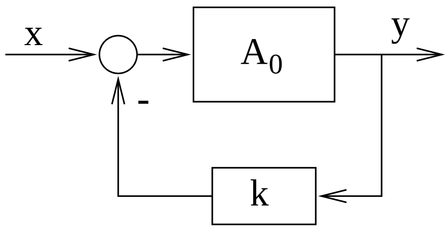
Natürlich gehören zur Angabe eines Klirrfaktors auch die Messbedingungen. Der Klirrfaktor ist bei konkreten Systemen frequenz- und amplitudenabhängig. Typische Werte sind: 33% für eine Verzerrung von einem Sinus- zu einem Rechtecksignal. Bei 10% ist Sprache gerade noch verständlich. Bei 1% sind sie in guter Musik gerade noch zu hören, während ordentliche HIFI Verstärker einen Klirrfaktor von $< 0.1\%$ haben.

Füttert man ein nichtlineares System mit mehreren hamonischen Frequenzen, so treten am Ausgang zusätzlich auch Summen und Differenzen von diesen Eingangsfrequenzen auf. Bei zwei Eingangsfrequenzen ω_1 und ω_2 entstehen allgemein alle Frequenzen der Form $|p \cdot \omega_1 \pm q \cdot \omega_2|$, mit $p, q = 0, 1, 2 \ldots n$ und $p + q \leq n$. Mit Hilfe von schmalbandigen Verstärkern kann man versuchen diese *Intermodulationsprodukte* loszukriegen, allerdings gibt es immer solche, die besonders nahe an den Nutzfrequenzen liegen.

Bei Systemen mit nur geringen Abweichungen von der Linearität wachsen die Intermdulationsfrequenzen k-ter Ordnung ($k = p + q$) mit u^k, wenn u die Amplitude des Eingangssignales ist! Der *Intermodulationsabstand* IMk ist das Verhältnis des Intermodulationsproduktes k-ter Ordnung zum Nutzsignal, das meist logarithmisch in dB angegeben wird. Er wird also mit zunehmender Aussteuerung des Signales kleiner (Skizze).

3.6 Rückkopplung

Für das Verständnis der im nächsten Kapitel zu besprechenden Operationsverstärkerschaltungen und auch für die Regeltechnik von zentraler Bedeutung sind rückgekoppelte Systeme. Man spricht allgemein von *Rückkopplung*, wenn das Ausgangssignal eines Systems – allenfalls verändert – auf den Eingang zurückgeführt wird, und dort mit dem ursprünglichen Eingangssignal verknüpft wird. Ist die Rückführung linear und die Verknüpfung am Eingang additiv spricht man von *Mitkopplung*, wird das zurückgeführte Signal subtrahiert, spricht man von *Gegenkopplung*.



Eine Beispiel einer Gegenkopplung haben wir bereits als Stromgegenkopplung bei der Emitterschaltung von Transistoren kennengelernt. Bezeichnen wir die maximale Verstärkung mit A_0, und subtrahieren wir den k-ten Teil des Ausgangssignales y vom Eingangssignal x, erhalten wir:

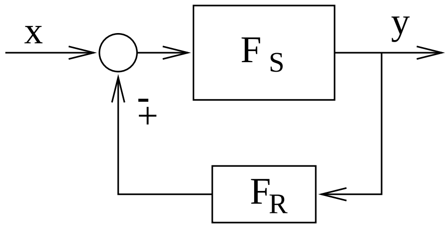
$$y = (x - k \cdot y) \cdot A_0 \tag{3.27}$$

Die Gesamtverstärkung A wird also

$$A := \frac{y}{x} = \frac{A_0}{1 + k \cdot A_0} \qquad \text{oder} \qquad \frac{1}{A} = \frac{1}{A_0} + k \tag{3.28}$$

A_0 heisst auch *offene Verstärkung* oder *open loop gain*, also die Verstärkung, die bei offenem Rückkopllungsweg wirkt. Die Grösse $k \cdot A_0$ heisst auch *Schleifenverstärkung* oder *loop gain*. Ist also die Schleifenverstärkung gross gegenüber 1, dann wird die Gesamtverstärkung gerade $1/k$. (Im Emitterfolgerbeispiel war das $-R_C/R_E$).

Die Operationsverstäkerschaltungen, die wir unten kennenlernen, bestehen aus solchen Rückkopplungsarten, wobei der Faktor k nicht nur durch Widerstände sondern im Prinzip mit Hilfe eines komplexen Netzwerkes oder sogar aktiven Elementen eine beliebige Übertragungsfunktion $F_R(s)$ als Rückführung realisiert werden kann.



Bezeichnen wir mit $F_S(s)$ die Übertragungsfunktion des Systems, so erhalten wir als gesamte Übertragungsfunktion eines Systems mit Rückführung:

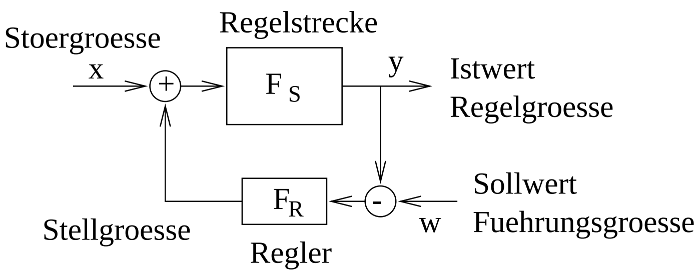
$$F(s) = \frac{F_S(s)}{1 \pm F_R(s) \cdot F_S(s)} \tag{3.29}$$

25.11.98

3.7 Regelkreise

(ab 14.12.98 dsikutiert)

Die besprochene Rückführung ist nichts anderes als ein Spezialfall eines Regelkreises. Ein solcher hat gegenüber der Rückführung noch einen zusätzlichen Eingang für die *Führungsgrösse* w (*Sollwert*). Unser Eingang x wird nun als *Störgrösse* bezeichnet, der Ausgang y wird zur *Regelgrösse* (*Istwert*).



Die Regelstrecke stellt dabei in der Regel ein physikalisches System dar, von dem eine bestimmte Eigenschaft (z.B. Temperatur) moeglichst genau auf einem bestimmten Wert gehalten werden soll.

Diese Grösse muss am Messort mit einem Messfühler gemessen werden, das Resultat dieser Messung stellt die Regelgrösse dar. Die *Regelabweichung* $y-w$ (oder *Regeldifferenz* $w-y$) liegt am Eingang des *Reglers*, der mit seiner Übertragungsfunktion F_R daraus die *Stellgrösse* (z.B. Heizleistung) erzeugt. Diese wirkt zusammen mit in der Regel unbekannten Störungen auf die *Regelstrecke* (z.B. Volumen mit Wärmekapazität) ein, die mit einer Übertragungsfunktion F_S die Regelgrösse y (z.B. Temperatur) verändert, usw.

Beachte, dass F_S die Änderung der Regelgrösse als Funktion der Stellgrösse beschreibt, diese Übertragungsfunktion der Regelstrecke enthält also auch das Übertragungsverhalten des Messfühlers. Schliesslich wirken Störungen nicht nur an der eingezeichneten Stelle, sondern im Prinzip überall auf den Regelkreis ein.

Man spricht von der *Störübertragunsfunktion* F_Z und der *Führungsübertragungsfunktion* F_w.

$$F_Z = \frac{\partial y}{\partial x} = \frac{F_S}{1+F_S F_R} \qquad F_w = \frac{\partial y}{\partial w} = \frac{F_S \cdot F_R}{1+F_S \cdot F_R} \tag{3.30}$$

Beide Funktionen haben den gleichen Nenner. $F_0 = F_R \cdot F_S$ heisst die *Kreisübertragungsfunktion.*

Im besonderen stellen wir fest, dass für frequenzunabhängige Regelkreise immer eine permanente Regelabweichung $w-y$ also eine Differenz zwischen Ist- und Sollwert bleibt: Ist nämlich $F_w(s) = K$ konstant (immer < 1), dann wird für eine beliebige Führungsfunktion $\mathrm{d}w(t)$ deren Laplacetransformierte mit K multipliziert. Die inverse Laplacetransformation liefert daraus $\mathrm{d}y(t)$. Wegen der Linearität der Laplacetransformation wird gerade $\mathrm{d}y(t) = K \cdot \mathrm{d}w(t)$.

Es bleibt dabei eine permanente Regelabweichung $\mathrm{d}w - \mathrm{d}y = (1-K)\cdot \mathrm{d}w$ bestehen, die umso kleiner wird, je grösser F_0 ist, und somit je näher K an 1 heran kommt.

3.7.1 Stabilität von Regelkreisen

Die Kreisübertragungsfunktion $F_0 = F_R \cdot F_S$ muss für die Diskussion der Regeleigenschaften und der Stabilitätsbetrachtung bekannt sein. Man misst sie, indem man den

Regelkreis beim Soll-/Istwert – Vergleich unterbricht, und dann Verstärkung und Phasengang des Systems misst, das heisst für harmonische Signale bei w, misst man y als Funktion der Frequenz.

Wird der Nenner zum Beispiel von F_x gleich null, ergibt sich ein Ausgangssignal (Änderung der Regelgrösse), selbst ohne dass ein Störung vorhanden ist. Der Regelkreis wird unstabil. Wie bei der Einführung der Laplacetransformation erkennen wir also, dass die Stabilität des Regelkreises durch die Pole der Übertragungsfunktionen definiert wird. Setzt man deren Nenner (der für F_w und F_x gleich ist!) gleich null, nennt man das die *charakteristische Gleichung* des Regelkreises:

$$1 + F_0(s) = 0 \tag{3.31}$$

Die Lösungen $s_k = \sigma_k + i\omega_k$ dieser Gleichung stellen gerade die komplexen Eigenkreisfrequenzen des Regelkreises dar. Aus der Theorie der Laplacetransformation (siehe die drei Regeln in Abschnitt 3.4) wissen wir, dass der Regelkreis mit der Frequenz ω_k schwingt, wenn $\sigma_k \geq 0$ ist, ansonsten bestimmen die s_k die Einschwingvorgänge nach Störsprüngen.

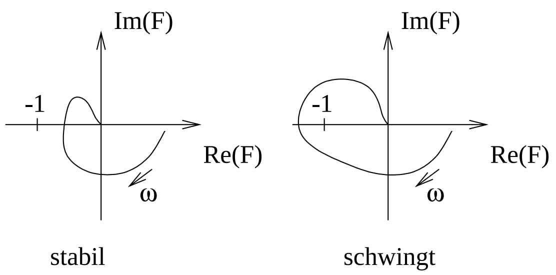
Ein Regelkreis ist also genau dann stabil, wenn alle $\sigma_k < 0$ sind. Wenn die charakteristische Gleichung höchstens 2-ten Grades ist:

$$a_0 + a_1 s + a_2 s^2 = 0 \tag{3.32}$$

dann sind offensichtlich genau dann die Realteile der Lösungen negativ, wenn alle Koeffizienten a_k positiv sind. Dies ist das *Hurwitzkriterium* für die Stabilität eines Regelkreises. Für charakteristische Gleichungen höherer Ordnung muss das Hurwitzkriterium durch kompliziertere Bedingungen ergänzt werden (siehe [Ebel78], Seite 33), die allerdings oft immer noch einfacher sind, als das Lösen der charakteristischen Gleichung.

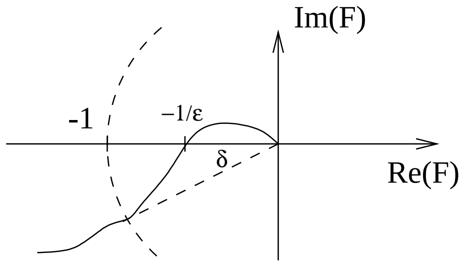
Statt die Nullstellen der charakteristischen Gleichung zu diskutieren, kann man die -1 Stellen von $F_0(s)$ untersuchen. Nach einem Satz aus der Funktionentheorie ist die Zahl der -1 Stellen in der rechten Halbebene von $F_0(s)$ gleich der Anzahl der Umfahrungen der Ortskurve von $F_0(i\omega)$ um -1. Für einen stabilen Regelkreis umschliesst sie also den Punkt -1 überhaupt nicht. (*Nyquistkriterium*).

Die folgenden zwei Beispiele zeigen solche Ortskurven, also die komplexen Werte der Funktion $F_0(i\omega)$.



Für $\omega \to \infty$ ist die Schleifenverstärkung in Praxis immer 0, die Ortskurven enden also im Nullpunkt der Darstellungen. Für Regelkreise mit Integratoranteilen $(1/i\omega)$, kommen die Ortskurven für $\omega = 0$ aus dem Unendlichen, in diesem Fall muss der Punkt -1 immer links liegengelassen werden.

Es reicht also die Kreisübertragungsfunktion als komplexen Frequenzgang (Amplituden- und Phasengang) numerisch zu kennen, um das Nyquistkriterium anwenden zu können.



Der Abstand der Ortskurve vom Punkt -1 sagt etwas über die Länge eines allfälligen Einschwingvorganges aus. Man definiert den *Phasenrand* δ, der die Phasenverschiebung bei Vertärkung $|F_0| = 1$ angibt, sowie den *Amplitudenrand* ϵ, der die Verstärkung (genauer Abschwächung $1/\epsilon$) bei der Phasenverschiebung 180° angibt.

Der Phasenrand wird auch *Phasenreserve* genannt. Er gibt direkt die Dämpfung des Einschwingvorganges an. Bei $\delta = 90°$ erhalten wir gerade den aperiodischen Grenzfall (kritische Dämpfung). Bei $\delta \approx 60°$ tritt bei der Sprungantwort ein Überschwingen von ca. 4% an, die Einstellzeit ist in diesem Fall besonders klein. (vergleich auch 3.4.1).

Normalerweise soll $\epsilon > 2.5$ und $\delta > 30°$ sein, um genügend stark gedämpfte Einschwingvorgänge zu bekommen.

3.7.2 P – Regler

Der *Proportionalregler* ist frequenzunabhängig. Der Ausgang y ist direkt proportional zum Eingang x:

$$y = K_P x \qquad F_R(s) = K_P \tag{3.33}$$

Die Sprungantwort besteht demnach ebenfalls aus einem Sprung. Proportionalregler sind stabil, sie erzeugen aber eine endliche permanente Regelabweichung, die umso kleiner wird, je grösser die Regelverstärkung ist.

3.7.3 PI – Regler

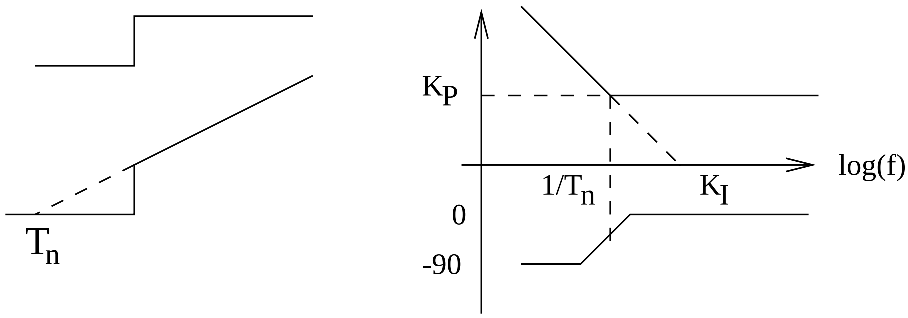
Will man die verbleibende Regelabweichung mit der Zeit zum Verschwinden bringen, muss die Kreisübertragungsfunktion einen integrierenden Anteil bekommen, der zum

Beispiel im Regler enthalten sein kann.

Der PI – Regler besteht aus einer Kombination von einem P und I System:

$$y = K_p x + K_I \int x \mathrm{d}t \qquad F_R(s) = K_P + \frac{K_I}{s} = K_P(1 + \frac{1}{T_n s}) \tag{3.34}$$

. $T_n = K_P/K_I$ heisst die Nachstellzeit des Reglers. Sprungantwort und Frequenzgang (Bodediagramm) eines idealen PI – Reglers sehen wie folgt aus:



Beachte, dass der PI Regler für kleine Frequenzen integriert, und für grosse proportional verstärkt, also gerade umgekehrt, wie ein Tiefpass. K_I heisst auch die *Durchtrittskreisfrequenz*, bei dieser ist die Verstärkung des I Anteils gerade 1. Der Übergang von dem integralen zum proportionalen Verhalten geschieht bei der Frequenz $1/T_n$, diese heisst deshalb *Eckfrequenz*. Bei sehr kleinen Frequenzen ergibt der ideale PI – Regler unendliche grosse Reaktionen, man muss diese in der Praxis künstlich beschränken, was nicht immer trivial ist. 14.12.98

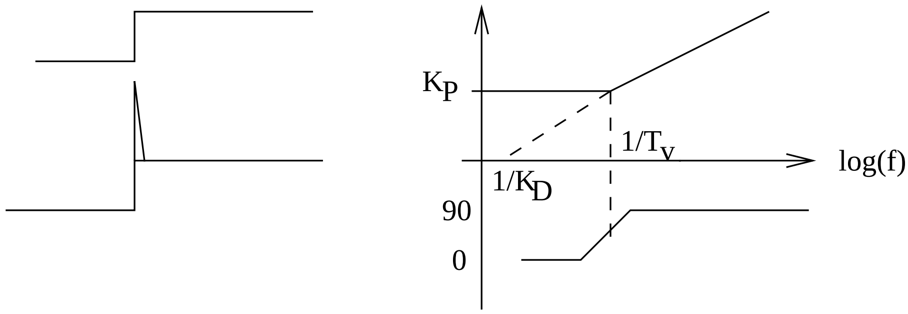
3.7.4 PD – Regler

Zur Erhöhung der Ansprechgeschwindigkeit eines Reglers auf Störungen oder Änderungen der Führungsgrösse kann dem P – Regler noch ein differenzierender Anteil dazugefügt werden:

$$y = K_P x + K_D \frac{\mathrm{d}x}{\mathrm{d}t} \qquad F_R(s) = K_P + K_D s = K_P(1 + T_v s) \tag{3.35}$$

$T_v = K_D/K_P$ heisst Vorhaltezeit.

Sprungantwort und Bodediagramm des idealen PD – Reglers sehen wie folgt aus.



Der ideale PD – Regler reagiert vorerst mit einer Deltafunktion. In der Praxis handelt es sich natürlich um einen beschränkten Ausschlag mit in der Regel exponentiellem Abfall. (siehe bei den Implementationen im nächsten Kapitel. Im Gegensatz zum Hochpass, differenziert der PD – Regler bei hohen Frequenzen, bei niedrigen dominiert der Proportionalanteil. Bei sehr hohen Frequenzen muss die Reaktion in der Praxis beschränkt werden.

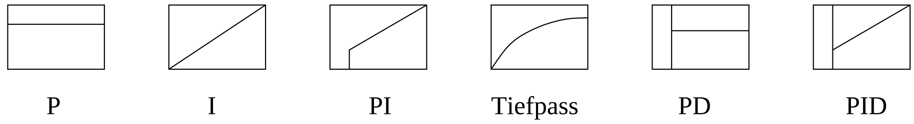
3.7.5 PID – Regler

Kombiniert man PI und PD zusammen, erhält man einen PID – Regler. Er hat den Vorteil, dass er auf Störungen schnell reagiert, und auf lange Frist die Regeldifferenz exakt auf null regelt. Seine Übertragunsfunktion lautet also:

$$F_R(s) = K_P(1 + \frac{1}{T_n s} + T_v s) \tag{3.36}$$

3.7.6 Reglersymbole

In der Regeltechnik werden die verschiedenen Übertragungssysteme durch rechteckige Symbole dargestellt, die im Innern eine vereinfachte Form der Sprungantwort skizziert haben:

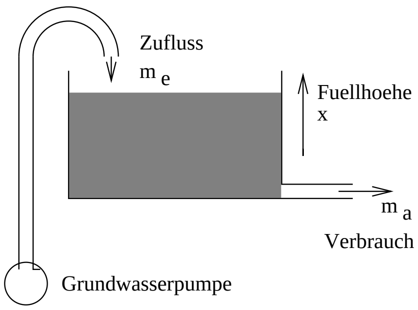


Für die Implementation dieser Regler siehe später.

3.7.7 Beispiel Reservoirregelung

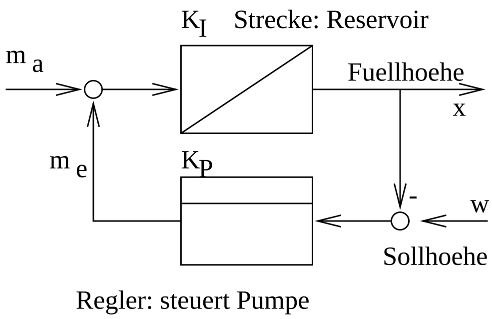
([Ebel78], page 47 ff. und 60 ff.)

Als Beispiel wollen wir die Dynamik der Regelung der Füllhöhe eines Wasserreservoirs diskutieren.



Das Reesrvoir wird durch eine Grundwasserpumpe mit dem Massenfluss m_e (Stellgrösse) gefüllt. Angeschlossene Kunden verbrauchen eine nicht vorhersehbare Menge m_a (Regelgrösse, Istwert) von Wasser. Ein Regler soll nun die Pumpe so steuern, dass eine möglichst konstante Füllhöhe x erreicht wird.

Als erstes müssern wir die Übertragungsfunktion der Regelstrecke studieren: Anschaulich erkennt man sofort, dass z.B. nach einem Sprung von m_a die Regelgrösse x konitnuierilich abnimmt. Es handelt sich also um ein integrierendes Verhalten: $F_S(s) = K_{IS}/s$.



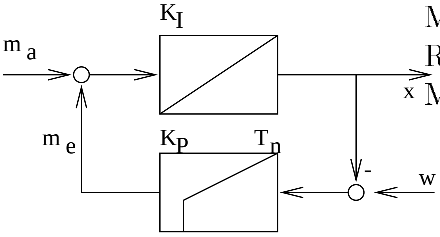
Im einfachsten Fall könnte man einen P–Regler einsetzen. Nebenstehendes Bild zeigt das Reservoir in der Symbolik der Regeltheorie. Die Regelstrecke umfasst dabei alle Elemente des Reservoirs vom Wasserfluss bis und mit des Messfühlers der Füllhöhe.

Die Störübertragunsfunktion wird

$$F_Z(s) = \frac{\partial x}{\partial m_a} = \frac{F_S(s)}{1 + F_R(s)F_s(s)} = \frac{K_{IS}/s}{1 + K_P K_{IS}/s} = \frac{1/K_P}{1 + T \cdot s} \tag{3.37}$$

wobei $T = 1/K_P K_{IS}$ und K_P die Regelkonstante des P – Reglers. Es handelt sich offenbar um ein Tiefpassverhalten. Die Sprungantwort wird proportinal zu $1 - e^{-t/T}$ (siehe bei der Tiefpassdiskussion). Es ergibt sich also eine permanente Regelabweichung von $\mathrm{d}m_a/K_P$.

Möchte man die Regelabweichung wegregeln, muss man einen I – Regler einsetzen. Ein reiner I – Regler würde aber zu einem Schwingen des Regelkreises führen, da wegen den beiden I Teilen die Phasenverschiebung von $F_0 = F_R F_S$ zu 180° und der Phasenrand somit zu 0 würde.



Man verwendet deshalb zweckermässigerweise einen PI – Regler mit Nachstellzeit T_n und Proportionalanteil K_P. Mit

$$F_S(s) = K_{IS}/s \qquad F_R(s) = K_P(1 + \frac{1}{T_n s}) \tag{3.38}$$

erhalten wir für die Störübertragungsfunktion

$$F_Z(s) = \frac{T_n}{K_P} \cdot \frac{s}{1 + T_n s + \frac{T_n}{K_P K_{IS}} s^2} \tag{3.39}$$

Der Nenner dieses Ausdrucks gibt uns durch Nullsetzen die charakterische Gleichung, die in diesem Falle quadratisch ist. Wir können deshalb das Hurwitzkriterium anwenden: Für einen stabilen Fall müssen alle Parameter psoitiv sein, insbesondere $T_n > 0$. Für $T_n \to 0$ (d.h. die Eckfrequenz des PI – Reglers geht gegen unendlich, d.h. kein P Anteil), wird das Hurwitzkriterium verletzt, der Regelkreis schwingt.

Setzen wir als Abkürzung

$$\omega_0^2 = \frac{K_P K_{IS}}{T_n} \qquad \theta = \frac{1}{2}\sqrt{K_P T_n K_{IS}} \tag{3.40}$$

so wird die charakterische Gleichung zu

$$1 + 2\frac{\theta}{\omega_0} s + \frac{1}{\omega_0^2} s^2 = 0 \tag{3.41}$$

und die Störübertragungsfunktion

$$F_Z(s) = \frac{K_{IS}}{\omega_0^2} \frac{s}{1 + 2\frac{\theta}{\omega_0} s + \frac{1}{\omega_0^2} s^2} \tag{3.42}$$

Um die Sprungantwort zu berechnen, müssen wir $F_Z(s)$ mit der lapalcetransformierten der Sprungfunktion, also $1/s$, multiplizieren, und das Resultat zurücktransformieren. Wir erhalten damit als Sprungantwort:

$$x(t) = \frac{K_{IS}}{\omega_0} \cdot \frac{1}{\sqrt{1-\theta^2}} e^{-\theta\omega_0 t} \cdot \sin(\omega_0 \sqrt{1-\theta^2} \cdot t) \tag{3.43}$$

In der Tat sehen wir, dass der Regelkreis für $\theta \to 0$ (kein P – Anteil) für immer schwingt. Dies entspricht dem Fall, wo die Lösungen der charakteristischen Gleichung rein imaginär sind und somit auf der imaginären Achse liegen, also nicht mehr im linken Halbraum. 16.12.98

Um eine optimale Dämpfung zu bekommen, muss $\theta \to 1$ gehen. Das bedeutet für die Regelparameter

$$T_n K_P = \frac{4}{K_{IS}} \tag{3.44}$$

Man erkennt hier auch wieder die wichtige Regel, dass für die Dimensionierung eines Reglers die Parmeter der Regelstrecke bekannt sein müssen.

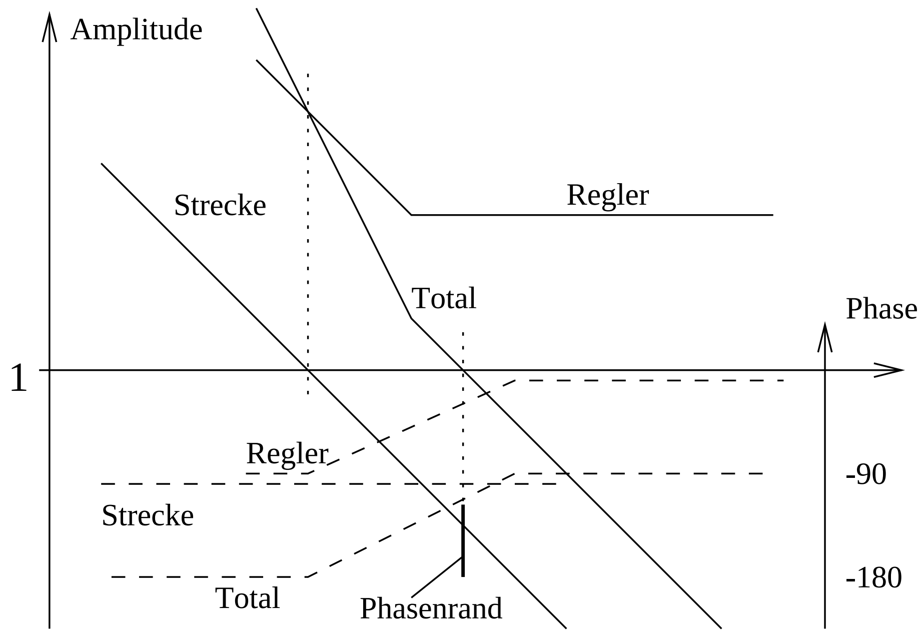
(Natürlich gibt es Fälle, wo die dynamischen Eigenschaften der Regelstrecke nicht zum vorneherein bekannt sind und sich evt. auch während dem Betrieb ändern können. Man kann aber natürlich mit einigem Aufwand auch Regelsysteme bauen, die ihre Eigendynamik laufend analysieren, und die Regelparamter entsprechend anpassen. Man spricht dann von *adaptiver Regelung.*)

Die charakterische Gleichung wird dann zu $(1 + s/\omega_0)^2 = 0$ mit der einzigen Lösung $s_0 = -\omega_0$. Wendet man auf diesen Fall unsere Methode wieder an, erhält man als Sprungantwort auf eine Störung

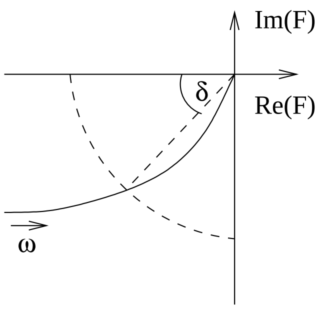
$$x(t) = K_{IS}\ \omega_0\ t\ e^{-\omega_0 t} \tag{3.45}$$

Die Regelabweichung wird also langfristig verschwinden. Wenn man ein leichtes Überschwingen in Kauf nimmt, dann wird die Regelabweichung schneller klein.

Im folgenden Bodediagramm sind Amplituden- und Phasengang der Regelstrecke und des Reglers sowie ihr Produkt ('total' = Kreisübertragunsfunktion) eingezeichnet. Beachte, dass der Phasenrand von der Eckfrequenz des PI – Reglers abhängt. Die optimale Einschwingdauer wird also durch die Wahl von T_n bestimmt. Um sicherzustellen, dass der Regler nicht schwingt, muss die Eckfrequenz so gewählt werden, dass der totale Amplitudengang die Achse mit der Steigung ≤ 1 schneidet.



Wie wollen auch das Nyquistkriterium anschauen:



Im nebenstehenden Bild ist die Ortskurve unseres Reglers dargestellt. Beachte, dass der Phasengang für sehr kleine Frequenzen fast 180° wird. Da dies jedoch weit weg vom Durchgang der Verstärkung durch die 0 dB Achse (Durchtrittsfrquenz) passiert, besteht keine Schwingungsgefahr.

Die Phasenreserve δ ist eingezeichnet. Der Amplitudenrand ist ∞, da die Phasenverschiebung von 180° nur asymptotisch erreicht wird.

Kapitel 4

Analoge Schaltungstechnik

4.1 Operationsverstärker: Aufbau und Kennwerte

(siehe [Kori98], S.342 bis 346)

Ein *Operationsverstärker* ist ein Verstärker mit sehr grosser Verstärkung. Er wird in der Regel mit Gegenkopplung durch ein meist passives Netzwerk betrieben, sodass die Verstärkung durch dessen Parameter bestimmt ist.

Ein Operationsverstärker hat zwei Eingänge V_+ und V_- und ein Ausgang V_o. Seine Verstärkung A_0 soll möglichst hoch sein, sie liegt bei kleinen Frequenzen im Bereich $10^4 \ldots 10^5$ und ist definert durch

$$V_o = A_0 \cdot V_d \qquad \text{mit} \quad V_d = V_+ - V_- \tag{4.1}$$

während der Eingangsstrom möglichst klein sein soll (kann durchaus im pA Bereich liegen). Üblicherweise haben Operationsverstärker eine bipolare Spannungsversorgung (meist ± 15V), der Ausgang kann alle Werte dazwischen annehmen.

Operationsverstärker sind in der Regel aus drei Teilstufen aufgebaut:

1. Die Eingangsstufe besteht aus einem Differenzverstärker (siehe Abschnitt 2.6.1) mit aktiver Last.
2. Eine Zwischenstufe (oft als Emitterschaltung ausgebildet) dient der weiteren Verstärkung und enthält die frequenzgangbestimmenden Elemente.
3. Die Endstufe besteht aus einem Gegentaktverstärker (siehe Abschnitt 2.6.3) im B–Betrieb (push–pull).

Operationsverstärker sind als integrierte Schaltungen in verschiedensten Bauformen (zum Teil auch mehrere in einem Gehäuse) erhältlich.

Kennwerte

Mit dem *Ausgangssteuerbereich* bezeichnet man den Wertebereich der Ausgangsspannung, seine Grenzen liegen in der Regel 0.2 bis 3 V über bzw. unterhalb der Versorgungsspannung (oft unsymmetrisch).

Der *Ausgangswiderstand* beträgt typisch 10…1000 Ω . (Beachte aber, dass dieser mit Hilfe der Gegenkopplung verändert werden kann).

Verschiedene Parameter, insbesondere die Verstärkung werden sowohl im Differenzbetrieb, als auch im Gleichtaktbetrieb angegeben. Die Differenzverstärkung A_d soll möglichst gross, die *Gleichtaktverstärkung* möglichst klein sein.

$$A_d = \frac{V_o}{V_+ - V_-} \qquad A_g = \frac{V_o}{\frac{1}{2}(V_+ + V_-)} \tag{4.2}$$

Unter *Gleichtaktunterdrückung* (*common mode rejection*) versteht man deren Verhältnis A_d/A_g.

Der *Eingangswiderstand* beträgt je nach Bauart bis zu $10^{12}\Omega$. Man unterschiedet zwischen Differenzeingangswiderstand und Gleichtakteingangswiderstand. Er vergrössert sich mit der Gegenkopplung noch.

Operationsverstärkung haben eine nicht vernachlässigbare *Offsetspannung*, das heisst eine endliche Eingangsdifferenzspannung, bei der die die Ausgangsspannung null wird. Sie beträgt in der Regel einige mV. Das kritische Grösse dabei ist deren Temperaturabhängigkeit, die *Offsetspannungsdrift*. Diese beträgt typisch einige μV/K.

Die Verstärkung nimmt bei hoher Frequenz ab. Der Effekt beginnt sich ab der *Grenzfrequenz* bemerkbar zu machen. Bei der *Transitfrequenz* f_T(auch *Verstärkungsbandbreiteprodukt* oder *unity gain freqeuncy* genannt) ist die Verstärkung gerade eins. Der Frequenzgang wird natürlich von der externen Beschaltung beeinflusst. Dabei ist besonders zu beachten, dass bei f_T die Phasenverschiebung sich von 180° signifikant unterscheidet (Phasenrand, siehe Abschnitt 3.7.1), da sonst der Vestärker schwingen wird.

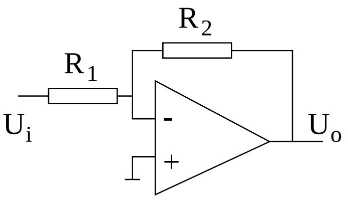
4.2 Anwendungen von Operationsverstärker

Für das elementare Verständnis der Operationsverstärkerschaltungen mit Gegenkopplung sind folgende beiden Regeln wichtig (gelten eigentlich nur für ideale Operationsverstärker):

1. Gegenkopplung versucht $V_d = V_+ - V_-$ auf null zu regeln.
2. Es fliesst kein Eingangsstrom

Die Gegenkopplung wurde im Abschnitt 3.6 formal diskutiert, und stellt die Grundlage aller folgenden Anwendungsschaltungen dar.

4.2.1 Operationsverstärker – Grundschaltungen



Beim invertierenden Verstärker wird der Eingang an V_- angeschlossen, V_+ liegt auf Masse. Der Gegenkopplungswiderstand R_2 bewirkt, dass der Regelkreis geschlossen ist. Wird V_- von Null verschieden, wird der Ausgang sich um die Verstärkung A_0 auf die Gegenseite bewegen, als Resultat wird die Spannung V_- auf null geregelt.

Wenn wir den Eingangsstrom vernachlässigen, fliesst durch die beiden Widerstände der gleiche Strom I, und es gilt nach dem Ohmschen Gesetz

$$I = \frac{U_i - V_-}{R_1} = \frac{V_- - U_o}{R_2} \tag{4.3}$$

Mit der Definition des open loop gains A_0 wird $U_o = A_0 \cdot V_-$. Man erhält dann für $A_0 \gg 1$ und $k = R_1/R_2$ die Übertragungsfunktion

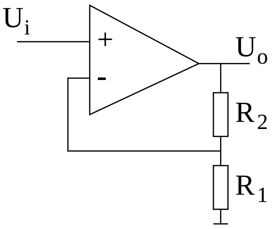
$$A = \frac{U_o}{U_i} = -\frac{A_0}{1 + k\ A_0} \tag{4.4}$$

Durch Vergleich mit 3.28 erkennt man, dass der Rückkopplungsfaktor offenbar gerade k und die Schleifenverstärkung $k\ A_0$ ist. Für grosse Schleifenverstärkung wird

$$A \approx -\frac{1}{k} = -\frac{R_2}{R_1} \qquad \text{für} \quad k\ A_0 \gg 1 \tag{4.5}$$

Im Falle einer grossen Schleifenverstärkung wird die Verstärkung also nur durch die Widerstände bestimmt, und ist unabhängig von der Temperatur-, Individualstreuungs- und Frequenzabhängigen open loop Verstärkung A_0.

Den Punkt V_- bezeichnet man mit *virtuelle Masse*, solange der Regelkreis funktioniert, ist $V_- \approx 0$. Die Eingangsimpedanz dieses Verstärkers wird somit gerade R_1, ist also relativ klein.



Der nichtinvertierender Verstärker hat im Gegensatz dazu einen sehr hohen Eingangswiderstand, der durch den Aufbau des Operationsverstärkers gegeben ist. Die Gegenkopplung erfolgt in diesem Fall über einen Spannungsteiler:

$$V_+ = U_i \; ; \quad V_- = U_o \frac{R_1}{R_2 + R_1} \quad \Rightarrow \quad A = \frac{A_0}{1 + k\,A_0} \; ; \quad k := \frac{R_1}{R_1 + R_2} \tag{4.6}$$

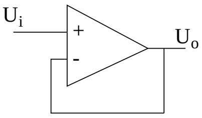
Für grosse Schleifenverstärkungen $k\,A_0$ erhalten wir wieder, dass die totale Verstärkung gerade

$$A \approx \frac{1}{k} = \frac{R_2}{R_1} + 1 \tag{4.7}$$

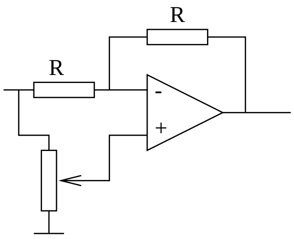
Diese Schaltung nennt man wegen ihres hohen Eingangswiderstandes auch *Elektrometerverstärker* 3.12.98

Spezialfälle der Grundschaltungen

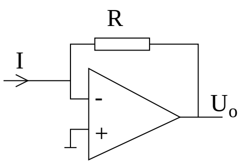
Die folgenden Schaltungen sind leichte Abwandlungen und Spezialfälle der Grundschaltungen:



Der *Impedanzwandler* ist ein nichtinvertierender Verstärker mit $k = 1$, also $A = 1$. Er dient zur Umwandlung einer hochohmigen Spannungsquelle in eine niederohmige.

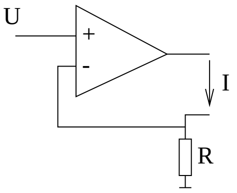


Das *bipolare Koeffizientenglied* erlaubt eine einstellbare Verstärkung über beide Vorzeichen: Steht das Potentiometer im oberen Anschlag ist sie +1 (Impedanzwandler), im unteren Anschlag -1 (invertierender Verstärker). In der Mittelstellung ist die Ausgangsamplitude gerade 0.



Der *Transimpedanzverstärker* wandelt einen Strom in eine Spannung um. Es wird $U_o = R \cdot I$. Der Eingangswiderstand ist R/A_0, also sehr klein, sodass eine Stromquelle als solche behandelt wird.

Transimpedanzverstärker kommen insbesondere als Verstärker für Sensoren, die einen Strom liefern, zum Einsatz.

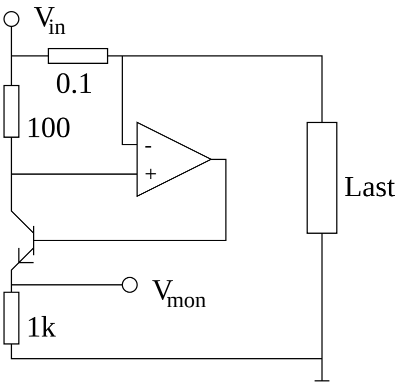


Der *Transadmittanzverstärker* oder *Transkonduktanzverstärker* (OTA, "operational transconductance amplifier") stellt eine spannungsgesteuerte Stromquelle dar: Der Aussgangswiderstand wird $R \cdot A_0$. Der Strom wird $I = U/R$.

Transkonduktanzverstärker werden besonders zum Beispiel für das Ansteuern von Leuchtdioden zur Signalübertragung oder für Kabeltreiber mit definierter Impedanz benützt.

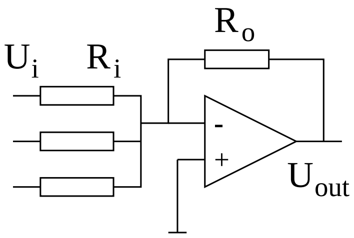
4.2.2 Weitere Operationsverstärkerschaltungen

Die folgenden Schaltungen stellen verschiedene Anwendungsbeispiele dar. Für das Verständnis halte man sich immer die beiden in der Einleitung erwähnten Regeln (Eingangsstrom=0, Eingangsdifferenz wird auf null geregelt) vor Augen.



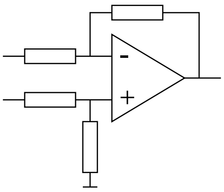
Diese Schaltung stellt einen *Strommonitor* dar (aus [Horo97], p 253). Die Spannung am Ausgang V_{mon} liefert eine Spannung, die proportional zum Strom in der Last ist, und zwar in der angegebenen Dimensionierung gerade 1 V/A. Der Operationsverstärker regelt dabei die Basisspannung des Transistors und damit dessen Kollektorstrom gerade so, dass der Spannungsabfall am 100 Ω Widerstand gleich dem am 0.1 Ω ist, also auf genau 1/1000 des Laststromes. Am 1 kΩ Widerstand fällt damit die gewünschte Spannung ab.

Der analoge *Addierer* besteht aus einem invertierenden Verstärker:



Da der V_- Eingang als virtuelle Masse fungiert, werden die Ströme, die die drei Eingangsspannungen U_i erzeugen, einfach aufaddiert. Durch Wahl verschiedener Widerstände R_i kann den Spannungen ein unterschiedliches Gewicht in der Summe gegeben werden:

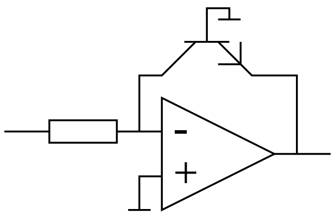
$$U_{out} = -R_o \cdot \sum_i \frac{U_i}{R_i} \tag{4.8}$$

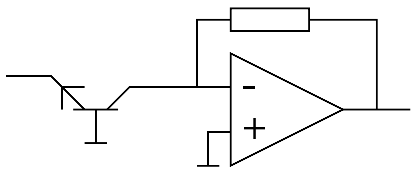


Falls alle 4 Widerstände genau gleich sind, stellt die nebenstehende Schaltung einen *Subtrahierer* dar. Versuchen Sie die Ausgangsspannung herzuleiten – auch für verschiedene Widerstände!

Nachteilig an der Subtrahierschaltung ist ihr relativ kleiner Eingangswiderstand. Schaltet man Elektrometerverstärker (siehe vorheriger Abschnitt) oder Impedanzwandler vor beide Eingänge des Subtrahierers, nennt man die entstehende Schaltung einen *Instrumentenverstärker*.

Diese folgenden beiden Schaltungen stellen einen *logarithmischer Verstärker* (links) und einen *exponentieller Verstärker* (rechts) dar.



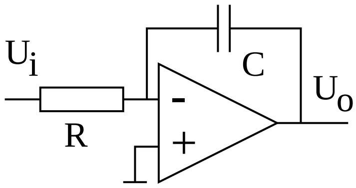


Dabei wird die exponentielle Form der Diodengleichung für die Basis-Emitter-Diode und die Linearität zwischen Basis- und Kollektorstrom ausgenützt:

$$I_C = I_0 e^{\frac{eU_{BE}}{kT}} \qquad \Rightarrow \qquad U_{BE} \sim \ln I_C \tag{4.9}$$

Diese Schaltungen werden für die Kompression und Dekompression von dynmischen Bereichen benützt: Schwanken die Messwerte über viele Dekaden, so ist es unmöglich, einen ADC zu bauen, der im ganzen Bereich hohe Genauigkeit hat. Mit einem vorgeschalteten Logarithmierer erreicht man, dass über den gesamten Messbereich der ADC den gleichen relativen Messfehler erzeugt.

Der *Integrator* basiert ebenfalls auf dem Konzept der virtuellen Masse:



Der durch den Widerstand fliessende Strom lädt den Kondensator auf, die Ausgangsspannung wird damit

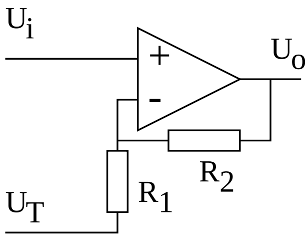
$$U_o = \frac{Q}{C} = \frac{1}{C}\int I \mathrm{d}t = \frac{1}{RC}\int U_i \mathrm{d}t \tag{4.10}$$

Man koennte also auch von einem Tiefpass mit unendlich kleiner Grenzfrequenz sprechen. Die Grenzferquenz wird in Praxis durch die Leerlaufverstärkung A_0 des verwedeten Operationsverstärkers definiert. Man vergegenwärtige sich das Bodediagramm.

In Praxis macht sich bei Integratoren die Eingangsoffsetspannung und der Eingangsruhestrom bemerkbar, die auch bei verschwindender Eingangsspannung ein langsames Driften der Ausgangsspannung bewirken. Die maximale Ausgangsspannung ("Anschlag") wird meist sehr schnell erreicht! Um dieses Problem zu Umgehen verwendet man häufig entweder Widerstände oder Schalter parallel zum Kondensator, die diesen kontrolliert entladen sollen.

Einen *Differentiator* bekommt man, wenn man in der obigen Schaltung R und C vertauscht. Diese Schaltung ist in Praxis aber sehr schwinganfällig, sodass man mit Vorteil noch einen zusätzlichen Eingangswiderstand in Serie schalten sollte (siehe [Tiet91], Seite 327).

Ein *Komparator* vergleicht eine analoge Spannung U_i mit einem Schwellwert (Threshold) U_T und soll eine positive oder negative Ausgangsspannung liefern, je nach dem das Eingangssignal grösser oder kleiner als der Schwellwert ist. Ein Operationsverstärker ohne jede Beschaltung mit Eingang an V_+ und Schwellwert an V_- angeschlossen, liefert offensichtlich gerade diese Funktionalität.



Oft möchte man eine *Hysteresis* im Schaltverhalten haben, was die nebenstehende *Schmitt–Trigger* Schaltung leisten kann. Die volle Breite der Hysterese wird gerade

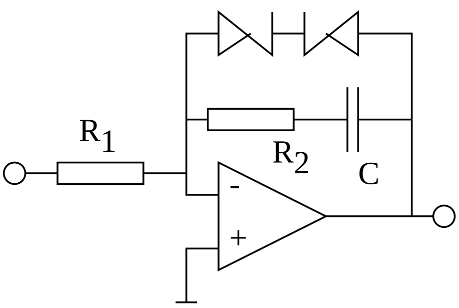
$$\Delta U_i = \frac{R_1}{R_1 + R_2} \cdot \Delta U_o\mathrm{max} \tag{4.11}$$

wobei $\Delta U_o\mathrm{max}$ die maximale Auslenkung von U_o bedeutet.

4.2.3 Realisierung von PI, PD und PID Reglern

(am 16.12.98 diskutiert)

Die im vorhergehenden Kapitel diskutierten Regler Typen können wir nun mit Operationsverstärkerschaltungen realisieren.

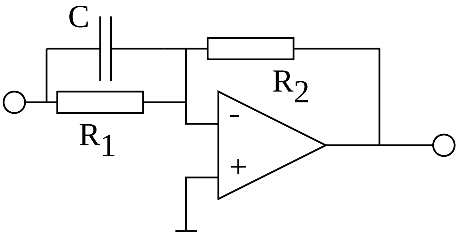


Der **PI – Regler** wird als Integrator realisiert. Zeige durch Grenzwertüberlegungen ($\omega \to \infty$, $\omega \to 0$), dass die Übertragunsfunktion dieses Reglers

$$-F_R(s) = \frac{R_2}{R_1} + \frac{1}{R_1 C s} \tag{4.12}$$

lautet. Identifiziere die PI – Regler – Konstanten K_P und T_n.

Die beiden Zenerdioden zeigen eine Möglichkeit auf, wie für ganz kleine Frequenzen die Ausgangsamplitude begrenzt werden kann, ohne dass sich eine Verschlechterung der Linearität des Reglers ergibt. Die Steilheit der Sprungantwort ist im übrigen durch die Slewtime des Operationsverstärkerausganges bestimmt.

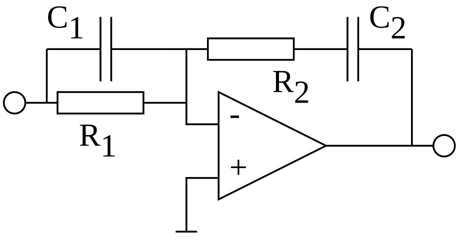


Der **PD – Regler**. Es gilt

$$-F_R(s) = \frac{R_2}{R_1} + R_2 C s \tag{4.13}$$

Die Eckfrequenz wird $1/R_1C$.

Die Deltafunktion der idealen Sprungantwort hat eine Anstiegszeit, die unter anderem von der Slew Time begrenzt ist, während der Abfall exponentiell erfolgt mit der Zeitkonstanten R_2C. Beachte, dass in dieser Zeit der Operationsverstärker nicht in seiner normalen Funktion arbeitet: Der - Eingang wird positiv, der Ausgang geht in den negativen Anschlag, die virtuelle Masse ist deshalb kurzfristig keine solche.



Der **PID – Regler** hat die Übertragungsfunktion

$$-F_R(s) = \frac{R_2}{R_1} + \frac{C_1}{C_2} + \frac{1}{R_1 C_2 s} + R_2 C_1 s \tag{4.14}$$

Diskutiere ausführlich das Bodediagramm, unter der Annahme, dass der Operationsverstärker den openloop gain A_0 hat, der ab der Frequenz $\omega_a \gg 1/R_2C_1$ mit Steigung -1 abfällt.

4.2.4 Anwendungen mit Dioden

Spitzenwertdetektor: [Tiet91], Seite 877. 7.12.98

Vollwegpräzisionsgleichrichter: [Tiet91], Seite 868.

4.2.5 Aktive Filter

(siehe [Kori98], Seite 360 ff.)

Mit Hilfe von Operationsverstärkern lassen sich fast beliebige Übertragungsfunktionen realisieren. Man spricht von Hoch-, Band- und Tiefpässen $n$ter Ordnung, wenn n energiespeichernde Elemente (C oder L) vorhanden sind. Gleichzeitig kommt n im Nennerpolynom als höchste Potenz vor und es gibt n Pole. Ihre Werte bestimmen die Form des Frequenzganges an der Grenzfrequenz: Konjugiert–komplexe Pole erzeugen einen positiven "overshooot" (Resonanzüberhöhung), reelle Pole einen langsamen Übergang vom Durchlassbereich in den Sperrbereich.

Aktive Filter höherer Ordnung lassen sich immer durch in Serie geschaltete Filter 2. Ordnung realisieren. Für diese wiederum kann man nun die Koeffizienten des Nennerpolynoms noch wählen. Optimiert man diese für möglichst gutes Rechteckübertragungsverhalten im Durchlassbereich, erhält man ein *Besselfilter*. Optimiert man auf möglichst glatten Frequenzgang erhält man ein *Butterworthfilter*. Will man eine maximalen Steilheit am Anfang des Sperrbereichs und nimmt man eine Welligkeit des Frequenzganges im Durchlassbereich in Kauf, verwendet man ein *Tschebyscheffilter*.

Tabellen der Koeffizienten, und die dazugehörigen Schaltungen findet man in [Kori98], Seite 363 ff., [Tiet91], Seite 391 ff. oder in [Best87].

4.2.6 Oszillatoren und PLL

Dreieck- Rechteckgenerator [Tiet91], p481 , auch mit 4 Dioden 9.12.98

ebenso gesteuert (VCO): [Horo97], p.240
Frequenz = $150 \times V_{in}/V_+$

'555

Skizze, Horowitz page 287 oben, aber rechte Logik als blackbox.

Einer der berühmtesten je hergestellten chips ist der '555 Timer. Er besteht aus nur aus zwei Komparatoren und etwas Schaltlogik. Er funktioniert mit einer einzigen Speisespannung (je nach typ von 1 bis 18 V). Die Bedeutung der drei Steuereingänge ist:

- $\overline{\textbf{reset}}$ (Pin 4) Ein negatives Signal setzt den Ausgang zurück auf LOW.
- $\overline{\textbf{trigger}}$ (Pin 2) Ein kurzes negatives Signal $< \frac{1}{3}V_+$ an diesem Eingang setzt den Ausgang auf HIGH.

- **threshold** (Pin 6) Ein kurzes positives Signal $> \frac{2}{3}V_+$ setzt den Ausgang auf LOW.

Neben dem OUTPUT (Pin 3) gibt es noch ein weiteres Ausgangssignal **discharge** (Pin 7). Es handelt sich um einen "open collector" Ausgang, der aktiv wird, wenn der output low ist. Schliesslich kann mit dem Eingang **reference** Pin 5, der Spannungsteiler beeinflusst werden.

Die einfachste Anwendung ist ein gewöhnlicher Oszillator

Zeichnung Horowitz, page 287 unten

Die Periode beträgt $T = \ln 2 \cdot (R_A + 2R_B) \cdot C$. Die Stabilität ist etwa 1% (Temperatur, Exemplarstreuung, Spannungsversorgung). Sie laufen etwa bis 3 MHz.

Der '555 wird auch zur Erzeugung von definierten Einzelpulsen (trigger als Eingang, zum Beispiel Power up reset) und anderes verwendet. Eine zeitlang war es bei den Elektroniksfreaks Mode in der Kaffeepause neue Anwendungsschaltungen für den '555 zu erfinden...

Bei obiger Schaltung ist die HIGH Zeit stets länger als die LOW Zeit. Für die Erzeugung sehr kurzer Pulse kann man die nebenstehende Schaltung verwenden, Horowitz page 288, die eine unabhängige LOW und HIGH Zeit erlaubt.

Verwendet man zum Laden des Kondensators eine Stromquelle, so kann man einen Sägezahn erzeugen (Horowitz, page 290). Mit einer steuerbaren Stromquelle kann man dann einen steuerbaren Oszillator bekommen.

Quartzoszillatoren

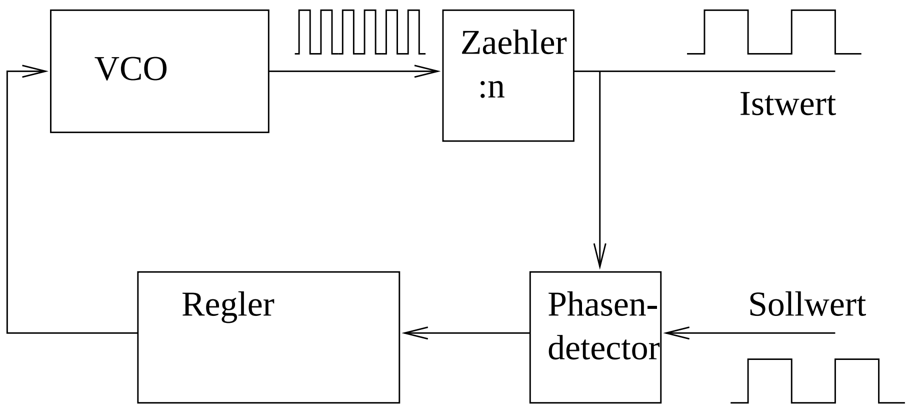
Quartz [Horo97], page 300, Parallel und Serieresonanz.

Nachlaufsynchronisation, PLL

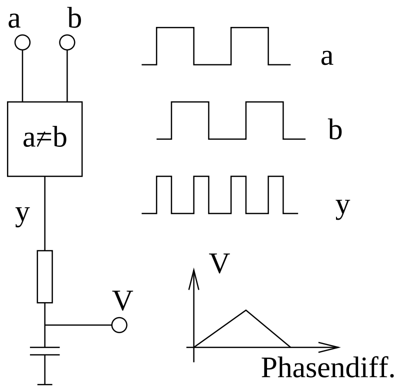
(siehe [Tiet91], p 954. [Horo97], page 641-653)

Der *phase locked loop* (PLL) ist ein wichtiger Anwendungsfall der Regeltheorie in der Nachrichtentechnik. Dabei möchte man die Frequenz f eines variablen Oszillators so einstellen, dass sie mit einer Bezugsfrequenz f_r übereinstimmt, uns zwar so genau, dass auch die Phasendifferenz der beiden Frequenzen genau konstant bleibt.

Eine typische Anwendung ist die Erzeugung einer Frequenz, die ein ganzzahliges Vielfaches einer externen Bezugsfrequenz ist.



Neben einem steuerbaren Oszillator VCO braucht man also einen Phasendetektor und einen Zähler, letztere werden wir in dem digitalen Teil diskutieren. Kritisch ist insbesondere die Dimensionierung des Reglers.



Der einfachste Phasendetektor besteht aus einem EX-OR Schaltung, die ein Signal produziert, wenn $a \neq b$ ist (siehe später in der Digitalelektronik). Dan anschliessende RC Glied mittelt diese Pulse, sodass eine Spannung entsteht, die proportional der Phasendifferenz ist. Beachte, dass es jeweils zwei Lösungen gibt, der Regelkreis wird aber nur auf einer Flanke stabil laufen, auf der anderen wird die Frequenz davonlaufen, bis die gute Flanke wieder erreicht wird.

Um zu verhindern, dass sich eine permanente Regelabweichung ergibt, möchte man als Regler gerne ein Integral - Regler verwenden. Die Regelstrecke stellt aber ebenfalls einen Integrator dar, da bei konstanter Frequenzdifferenz die Phase allmählich zunimmt. Zusammen ergäbe sich somit eine Phasenverschiebung innerhalb des Regelkreises von 180° , was zu einem schwingenden Regelkreis führen würde.

Man verwendet deshalb einen PI-Regler mit einer Eckfrequenz f_1, die unterhalb der Stelle f_2 liegt, wo die Kreis - Verstärkung 1 ist. (vernünftig ist 1/4 bis 1/10 davon, z.B. $f_2 = 2$ Hz, $f_1 = 0.5$ Hz).

Kapitel 5

Sensoren

Sensoren oder *Messfühler* dienen der Umformung der zu messenden physikalischen Grösse in eine elektrische Spannung, Strom oder Widerstandsänderung, die von Elektronik und Datenacquisitionssystemen weiterverarbeitet werden können.

Messfühler haben genauso wie alle anderen Systeme eine Übertragunsfunktion $F_M(s)$, die sowohl vom Fühler, wie seiner Ankopplung an das zu untersuchende System abhängt. Zum Beispiel hat ein Temperaturfühler ungefähr das Verhalten eines Tiefpasses, seine charakterische Zeit hängt von dem Umgebunsmedium (z.B. Wasser oder Luft) ab.

Früher wurde gefordert, dass der Zusammenhang zwischen der physikalischen Grösse und der abgegebenen Spannung möglichst linear sein soll. Dies Forderung ist in Hinblick auf die durch nachgeschaltete Prozessoren heute leicht automatisch durchführbare Linarisierung abgelöst worden durch die Forderung nach Stabilität und Reproduzierbarkeit der Übertragunsfunktion.

5.1 Licht

Strahlungsdetektoren beruhen grundsätzlich auf dem *Photoeffekt*, wobei man zwischen *äusserem* und *innerem* Photoeffekt unterschieden wird.

Beim ersteren wird ein Teil der vom Photon auf das Elektron übertragenen Energie dazu benutzt, um die Austrittsarbeit zu überwinden, das Elektron kann nun mit einem separaten Detektor nachgewiesen werden. Die Auetrittsarbeit stellt eine untere Schwellenenergie für den Nachweis von Licht dar.

Bei inneren Photoeffekt wird die Energie dazu verwendet, um zum Beispiel in einem Halbleiter den Unterschied zwischen Valenz- und Leitungsband zu überwinden und so zusätzliche Leitungsträger zu erzeugen. In diesem Fall ist die Energiedifferenz zwischen Loch- und Elektronzustand für die untere Detektionsschwelle relevant. Mit Hilfe von Dotierungen kann diese sehr niedrig gemacht werden.

5.1.1 Photovervielfacher

Photovervielfacher oder *Photomultiplier* funktionieren nach dem Prinzip des äusseren Photoeffekts. Elektronen, die aus der Photokathode emittiert werden, werden durch eine Anzahl von weiteren Elektroden (Dynoden) beschleunigt, wobei sie aus den Dynoden Sekundärelektronen auslösen (typisch 6 bis 10 Stück). Das ganze System muss sich im Vakuum befinden.

Die *Quanteneffizienz* η, also der Bruchteil der Photonen, die an der Photokathode ein Elektron erzeugen, beträgt in der Regel weniger als 20%. Die untere Energiegrenze wird durch die Austrittsarbeit des Materials (Cäsium oder Cäsiumlegierungen) gegeben, die obere durch die Durchsichtigkeit des Vakuumfensters für die Strahlung.

Die Gesamtverstärkung des Vervielfachers beträgt ca. 10^6 bis 10^8 hängt aber über die Zahl der Sekudärelektronenemission sehr stark vom elektrischen Feld ab, sodass die Spannung sehr gut konstant gehalten werden muss.

Photovervielfacher werden mit Vorteil dort eingesetzt, wo

- schnelle und zeitgenaue Signale benötigt werden (typische Verzögerungszeit ca 30 ns, sehr stabil, typische Anstiegszeit 1 ns).
- bestes SIgnal/Rausch verhältnis gefordert ist, insbesondere wenn eine rauschfreie Detektion von einzelnen Photonen nötig ist.
- hohe Homogenität der Empfindlichkeit auf der Dektektionsfläche.

Das Langzeitverhalten der Verstärkung von PM's ist allerdings nicht besonders gut, und nicht vorherzusagen, sie eignen sich also nicht für genaue Amplitudenmessungen. Sie haben eine schlechte Quanteneffizienz. Für Systeme mit vielen Kanälen sind sie ausserdem zu gross, zu teuer und zu unhandlich.

5.1.2 Photodioden

Photodioden bestehen aus einer *pn* Halbleitergrenzschicht, die in der Regel in der Sperrrichtung betrieben wird. Man verwendet zum Beispiel leicht dotiertes n-Basismaterial und bringt eine sehr stark dotierte, sehr dünne, durchsichtige p-Schicht darüber an. Das ergibt eine dicke Sperrschicht, die sich vor allem auf der n-Seite befindet. Die darin zu ELektron–Loch Paaren konvertierenden Photonen erzeugen einen zum normalen Sperrstrom (*Dunkelstrom*) zusätzlichen Photostrom I_{ph}.

Das Anlegen einer Sperrspannung erhöht die Dicke der Sperrschicht zusätzlich, erhöht also die Absorbtionswahrscheinlichkeit und somit die Effizienz. Ausserdem wird dadurch die Kapazität vermindert, sodass die Photodiode schneller anspricht und das Rauschen des anschliessenden Verstärker sich weniger auswirkt. Meistens wird die Effizienz dadurch verbessert, indem man p.i.n. Dioden verwendet, sodass die gesamte intrinsische Schicht als Konverter genutzt werden kann.

Sei Φ der Energiefluss des Lichtes, η die Quanteneffizienz, dann wird

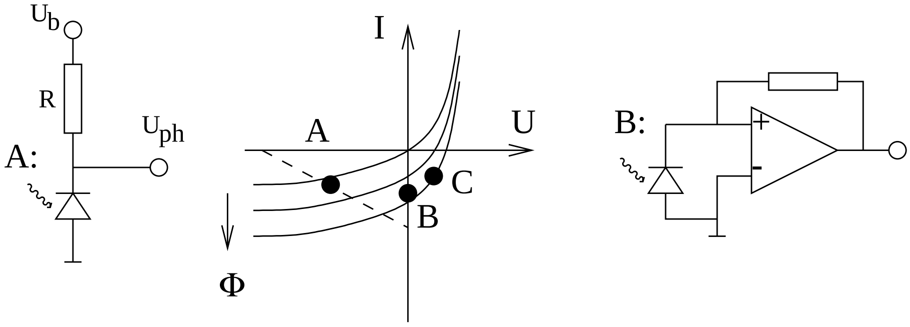
$$I_{ph} = e \cdot \eta \frac{\Phi}{h\nu} \tag{5.1}$$

Der Photostrom nimmt mit der Wellenlänge also zu, bis zu einer Grenzwellenlänge λ_{max}, die durch die Schwellenergie für Erzeugung von ELektron–Loch Paaren gegeben ist. Dieser Photostrom addiert sich zum normalen aus der Diodengleichung, der Gesamtstrom wird

$$I = I_0(e^{\frac{eU}{kT}} - 1) - \frac{e\eta\Phi}{h\nu} \tag{5.2}$$

Die Kennlinie verschiebt sich also in Stromrichtung parallel.

Je nach Anwendung unterscheidet man im wesentlichen drei verschiedene Beschaltungen. Für höchste Empfindlichkeit und Linearität auch bei ganz kleinen Intensitäten wird der Arbeitsbereich A mit Vorspannung U_b gewählt. Allerdings ergibt sich ein Schrottrauschen des Dunkelstromes. Je grösser U_b desto kleiner wird aber die Kapazität und also desto schnellere Ansteigszeit des Signales kann erreicht werden.



Der Kurzschlussbetrieb B eignet sich vor allem für genaue und reproduzierbare Messungen der Lichtintesität mit sehr guter Linearität, die zweckmässigerweise mit dem skizzierten Transimpedanzwandler realisiert wird. Da die Spannung nun verschwindet, wird die Kapazität nicht geladen, es ergibt sich eine schnellere Ansprechgeschwindigkeit, allerdings verstärkt die Kapazität nun das Eingangsrauschen des Operationsverstärkers. Diese Schaltung eignet sich weniger für kleine Lichtintensitäten.

Soll die Photodiode als Solarzelle Energie liefern, wird sie im Punkt C im sogenannten photovoltaischen Mode betrieben. Hier ist sie allerdings langsam in der Reaktion und stark nichtlinear in der Lichtausbeute. Dafür wird wegen dem fehlenden Dunkelstrom in dieser Schaltung auch die höchste Empfindlichkeit für schwaches Licht erreicht.

5.1.3 andere Halbleiterphotodetektoren

Mit Hilfe von *Phototransistoren* kann das Photosignal verstärkt werden. Dabei wird ein solcher Transistor in Emitterschaltung betrieben. Der Basisanschluss bleibt offen, in der gesperrten Basis–Kollektordiode werden durch das Licht zusätzliche Leitungsträger erzeugt, die in die Basis gelangen, und so einen Basis–Emitterstrom erzeugen, der dann durch die Stromverstärkung β entsprechend verstärkt wird. Wegen der Millerkapazität können so nur Frequenzen bis etwa 100 kHz erreicht werden.

Avelanche Photodioden sind prinzipiell aufgebaut wie p.i.n. Dioden, erlauben aber so hohe SPerrspannungen, dass im Inneren Stossionisation und mit dem entstehenden Lawineneffekt eine Signalverstärkung bis zu 10^4 möglich wird.

Photoleiter oder *Photowiderstände* bestehen aus einem einfachen Halbleiter. Lichteinstrahlund ändert die Leitfähigkeit infolge Ladungsträgererzeugung. Man misst die Widerstandsänderung, die sehr grosse Faktoren annehmen kann. Z.B. variiert der Widerstand eines CdS Photowiderstandes von 100 Ohm (hell) bis 10^7 Ohm (dunkel). CdS hat seine maximale Empfindlichkeit im sichtbaren Bereich, Silizium im infraroten Bereich.

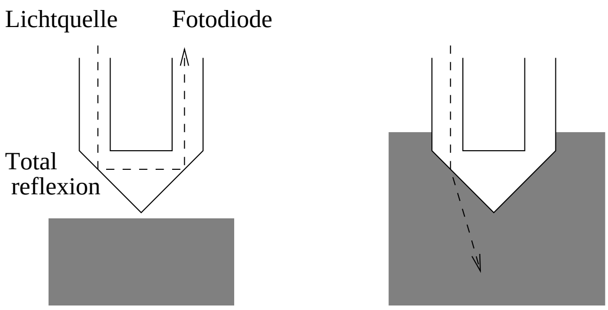
Solarzellen sind Photodioden, die im Punkt C betrieben werden. Sie werden optimiert für maximalen Wirkungsgrad. Die Leerlaufspannung U_l beträgt etwa 0.5 V (sinkt bei höherer Temperatur), der Kurzschlusstrom I_k ist proportional zum einfallenden Licht. Bei einem bestimmten Arbeitspunkt U_m, I_m erhält man die maximale Leistung $P_m = U_m I_m$, deren Verhältnis zu $U_l I_k$ heisst der Füllfaktor, der etwa 0.75 beträgt.

Man muss darauf achten, dass die Solarzellen nicht durch Infrarotabsorbtion erwärmt werden. Eine Kombination von Antireflexionsschichten und geschickte angeordnete Reflektoren an der Unterseite erhöhen den Wirkungsgrad.

5.2 Anwendungen von Lichtdetektoren

Positionen werden meist mit durch Spiegel abgelenkte Lichtstrahlen gemessen, die durch Fotodioden detektiert werden. Dabei wird entweder die Laufzeit des Signales (zum Beispiel Infratrotentfernungsmessung) oder die Position des Lichtstrahles durch mehere Fotodioden ermittelt. Im folgenden exemplarisch ein paar Anwendungen dieser Prinzipien:

5.2.1 Füllstandsanzeiger

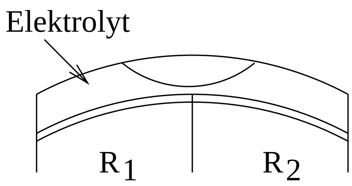


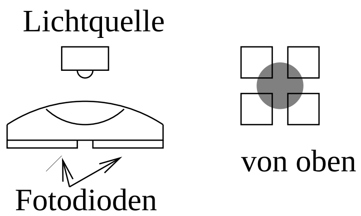
Eine Lichtquelle wird in einem Glaskörper mit geeigneter Geometrie totalreflektiert und von einer Fotodiode detektiert. Steigt die Flüssigkeit, so wird die Totalreflexion aufgehoben, und die Fotodiode wird dunkel, was einen erhöhten Flüssigkeitsspiegel signalisiert.

5.2.2 Drehwinkel

Drehwinkelmesser bestehen aus Scheiben, die ein geeignetes Strichmuster aufgedruckt haben. Man misst das davon reflektierte Licht, und zählt so die Streifen. Damit ist eine Winkeländerung messbar. Durch zwei Lichtquellen, die um eine halbe Phase des Strichmusters verschoben montiert sind, lässt sich ausserdem die Drehrichtung detektieren. Will man die Drehposition absolut messen, kann man verschiedene Muster parallel verwenden, die verschiedenen Bits entsprechen.

5.2.3 Neigunswinkelmesser





Das obere Bild zeigt ein gebogenes mit Elektrolyt gefülltes Röhrchen, wie eine Wasserwaage. Man misst das Verhältnis des Widerstandes R_1 und R_2, am besten mit einer Messbrücke. Man muss Wechselspannung verwenden sonst gibts elektrochemische Prozesse, die den Sensor zerstören.
Das untere Bild zeigt eine verbesserte Version. Die Lichtquelle wird durch die Luftblase auf die 4 Photodioden fokussiert. Neigt sich der Sensor, verschiebt sich der Lichfleck und die 4 Fotodioden sehen eine veränderte relative Lichtstärke, deren Auswertung Betrag und Richtung der Neigung ergeben kann.

5.3 Temperatur

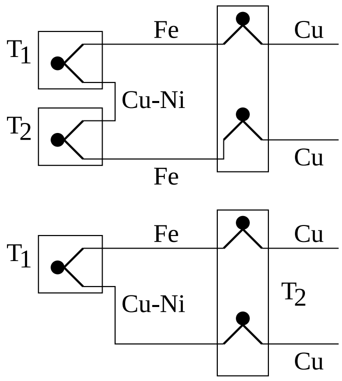
Metalle haben einen positiven Temperaturkoefizienten. Besonders gut linear sind *Platinwiderstände*, deren Widerstand

$$R = R_0(1 + 3.91 \cdot 10^{-3} \cdot T) \tag{5.3}$$

wobei die Temperatur in Celsius eingesetzt werden muss. *Platintemperaturfühler* werden zum Beispiel mit *Pt100* bezeichnet, wobei die Zahl 100 den Widerstand in Ω bei 0° C angibt. Sie lassen sich im Bereich von -200 bis + 850 C einsetzen. Da die Widerstandsänderung klein ist, wird normalerweise eine Messbrücke in 4-Draht Methode (Kelvinschaltung) verwendet, da sonst die Temperaturabhängigkeit der Messkabel die Messung verfälschen würde.

Halbleiterwiderstände haben normalerweise negative Temperaturkoeffizienten, da entsprechend der Fermifunktion bei höheren Temperaturen mehr Leitungsträger vorhanden sind. Man spricht von *NTC*. Ihr Widerstand variiert ziemlich stark, typisch 2% pro Grad. *PTC*'s haben einen positiven Temperaturkoeffizienten ähnlicher Grösse. PTC's kÖnnen direkt zur Leistungsbegrenzung bei zu starker Erwärmung in die Stromversorgung geschaltet werden.

Thermoelemente beruhen auf dem *Seebeckeffekt*, der die temperaturabhängige Spannung (*Thermospannung*) am Übergang zweier verschiedener Metalle beschreibt. Typische Temperaturkoeffizienten sind sehr klein z.B. 50 μV/K für Eisen–Konstantan. Sie eignen sich aber für sehr hohe Temperaturen, z.B. Wolfram–Rhenium bis 2800° C.



In der Praxis wird immer die Temperaturdifferenz zweier übergänge gemessen. Das obere Bild zeigt zwei gleiche Eisen – Konstantan – Fühler mit den Temperaturen T_1 und T_2. Dabei sind die beiden Übergänge zwischen dem Eisendraht und den Kupfermesskabeln genau auf der gleichen Temperatur zu halten! Das untere Bild zeigt eine vereinfachte Anordnung, wo der Übergangsblock vom Fühler zum Kupferkabel gleichzeitig die Referenztemperaturmessung T_2 darstellt.

5.4 Druck und Kräfte

5.4.1 Piezoeffekt

Der *piezoelektrische Effekt* kommt durch die relative Verschiebung von ungleich geladenen Atomschichten innerhalb eines Kristalles zustande, die eine elektrische Spannung erzeugt. Der bekanntest Piezomesskristall ist Quartz (SiO_2), aber auch künstliche Keramikarten zeigen Piezoeffekt.

Piezodruckfühler gibt es in allen Formen, zum Beispiel auch als Unterlagscheiben.

Die Messung stellt hohe Anforderung an den Eingangsverstärker und das Anschlusskabel. So bewirkt zum Beispiel bereits eine kleine Formänderung des Kabels (verschieben) eine fehlerhafte Messung, indem die Formänderung eine Kapazitätsänderung und darum eine Spannungsänderung bewirkt! Der Eingangsverstärker wird mit einem MOSFET versehen, um die Eingangsimpendanz hoch zu halten. Die Bandbreite wird möglichst nur für kleine Frequenzen ausgelegt, um nicht zu viel Rauschen zu bekommen.

Eine Anwendung davon stellen die Flussmeeungen dar, die

5.4.2 Dehnungsmessstreifen

Ein Draht, den man in die Länge zieht, wird nicht nur länger, sondern auch dünner, sein Widerstand nimmt also zu

$$R = \rho\frac{l}{A} = \rho\frac{l^2}{V} \quad \Rightarrow \quad \frac{\mathrm{d}R}{R} = S \cdot \frac{\mathrm{d}l}{l = S \cdot \frac{\sigma}{E}} \tag{5.4}$$

und zwar offenbar linear mit der Dehnung. Befestigt man einen solchen Draht in Serpentinen auf die Oberfläche des Körpers, dessen Dehnverhalten uns interessiert, wird der Effekt verstärkt, man spricht von einem Dehnungsmessstreifen. Sie werden auch als Kraftmesser eingesetzt.

Eine spezielle Anwendung dazu sind Membrandruckmesser. Dabei werden die einespannten Membranen an meherer Stellen mit Dehnungsmesstreifen versehen. Je nach Postion des Streifens entsteht bei hÖherem Druck, also bei hÖherer Auslenkung entweder eine Vergrösserung oder eine Verkleinerung des Widerstandswertes. Gemessen wird auch wieder mit einer Messbrücke.

Kapitel 6

Elemente der Digitalelektronik

6.1 Digitale Systeme und bits

Die *Digitalelektronik* befasst sich mit Nachrichtenübertragung und -verarbeitung von *binären Zuständen*, auch *digitale Systeme* S genannt. S ist eine endliche Menge von Elementen a_i, die auch *bit* genannt werden (von englisch *binary digit*):

$$S = \{a_1, a_2, a_3, \dots a_n\} \qquad a_i \in \{0, 1\} \tag{6.1}$$

Ein System aus n Elementen heisst $n - bit$ *System*. Die bits können also die Werte (auch Zustände genannt) 1 oder 0 annehmen, manchmal verwendet man auch T und F für *True* und *False*.

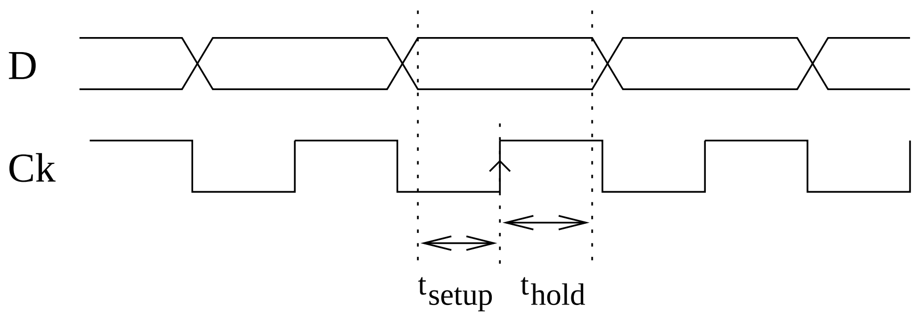
Einzelne Bits können aus analogen Signalen durch einen Schmitt–Trigger (siehe Abschnitt 4.2.2) direkt erzeugt werden.

Um mit solchen Systemen praktisch arbeiten zu können, muss eine Abbildungsvorschrift in eine technische Realisierung definiert werden: Es gibt hauptsächlich vier verschiedene **statische** Normen, die solche Abbildungen definieren. Dabei ist jeweils ein Zustand L (*Low*) und H (*High*) bechrieben. Man spricht von positiver Logik, wenn dem Zustand H der Wert T bzw. 1 zugeordnet ist, und dem Zustand L der Wert F bzw. 0. Ist die Zurdnung umgekehrt spricht man von negativer Logik.

	Low	High	
TTL	0.2 V (< 0.8 V)	3.5 V (> 2.0V)	
CMOS	0 V ($< \frac{1}{3}V_b$)	V_b ($> \frac{2}{3}V_b$)	V_b = Betriebsspannung
ECL	−1.7V (< −1.5V)	−0.9V (> −1.1V)	
Schaltkontakt	geschlossen	offen	

Natürlich gehören zu diesen Definitionen eigentlich auch die maximalen Innenwiderstände oder minimalen Stromstärken der Quellen, sowie die minimalen Innenwiderstände der Verbraucher. CMOS–HC Ausgänge können in der Regel TTL Bausteine treiben, während das umgekehrte (von TTL nach HC, auch von NMOS nach HC) mit einem 4.7kΩ pullup Widerstand nach +5V möglich ist.

Diese statischen Abbildungen erfordern für jedes bit eine eigene Leitung auf dem der Zustand dargestellt ist. Ändern sich die Daten im Laufe der Zeit, so werden sie zeitlich hintereinander, also **seriell** dargestellt. Werden mehrere verschiedene bits auf einer Leitung zeitlich hintereinander abgebildet, spricht man von einer *multiplexten* Darstellung. In diesen Fällen ist neben den Datenleitungen eine zusätzliche *Taktleitung* oder *clock* nötig, die mittels eines Rechtecksignales die Zeitpunkte definiert, bei denen die Daten sicher ihre richtigen Werte L oder H angenommen haben:



Die Abbildung zeigt ein Beispiel eines *timing diagrams*, wie es für die Beschreibung aller Funktionalitäten in der digitalen Elektronik verwendet wird. Die obere Zeile stellt eines der Datensignale D dar, das jeweils im L oder H Zustand sein kann. Die untere Zeile stellt das Taktsignal Ck dar, von dem es in der Regel eins pro System gibt. Es stellt ebenfalls ein logisches Signal dar, das abwechslungsweise L und H ist. Man bezeichnet die *aktive Flanke* mit einem Pfeil. Zwischen der *setup time* vor der aktiven Flanke und der *hold time* nach derselben sind die Werte auf den Datenleitungen stabil.

Zeitliche Abfolgen von solchen Bitwerten stellen Signale im Sinne von Kapitel 3.1 dar, ihre Übertragung unterliegt allen Regeln der analogen Elektronik. In der Praxis zeigt sich, dass vor allem bei grösseren Systemen die analogen Aspekte der digitalen Elektronik eine wesentliche Rolle spielen: Rauschen, Pickup, Übersprechen (X-talk), Reflexionen. Diese Dinge werden im Kapitel 8 diskutiert.

6.2 algebraische Grundlagen der Logik

Die erwähnten Systeme $S = \{a_1, a_2, a_3, \ldots a_n\}, a_i \in \{0, 1\}$ bilden zusammen mit den Operationen AND (*Konjunktion*), OR (*Disjunktion*) und NOT (*Negation*) eine **Algebra**, die *Boole'sche Algebra* oder die *Schaltalgebra* genannt. Die Wertetabellen für die drei Operationen mit Eingangswerten a und b, sowie dem Resultatwert y lauten:

AND

a	b	y
0	0	0
0	1	0
1	0	0
1	1	1

OR

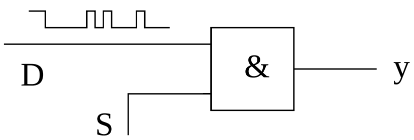
a	b	y
0	0	0
0	1	1
1	0	1
1	1	1

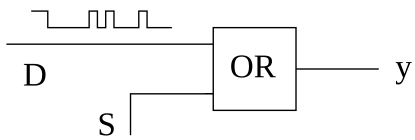
NOT

a	y
0	1
1	0

Man schreibt auch für AND: $y = ab$ oder $y = a \wedge b$. Für OR: $y = a + b$ oder $y = a \vee b$. Für NOT: $y = \bar{a}$ oder $y = \neg a$.

Das Einselement für die OR Operation ist die 0, dasjenige für die AND Operation die 1: $a + 0 = a \quad a \cdot 1 = a$. Anwendung: elektronischer Schalter für Datensignale:





Im linken Fall werden die Daten druchgelassen, das heisst $y = D$, falls $S = 1$ ist, andernfalls ist $y = 0$. Im rechten Fall werden die Daten durchgelassen, falls $S = 0$ ist, andernfalls ist $y = 1$.

Praktische Digitalbausteine, die die Grundfunktionen realisieren, heissen *Gatter* oder *gates*. Ein integrierter Baustein hat normalerweise mehrere Gatter im gleichen Gehäuse, oft sind es vier.

Man beobachtet, dass durch Anwendung der Negation auf die eine der beiden Wertetabelle für AND oder OR die jeweils andere Tabelle entsteht (**Dualität**).

$$\overline{\bar{a} + \bar{b}} = ab \tag{6.2}$$

Das bedeutet, dass die beiden Operationen durch Umkehrung der Logik (positive Logik, bzw. negative Logik) ineinander übergeführt werden können. Von dieser Möglichkeit macht man in praktischen Schaltungen oft Gebrauch, da damit die verschiedenen gates optimal genutzt werden können, allerdings wird naturgemäss die Übersichtlichkeit der Schaltung nicht gerade verbessert.

Es gelten folgende **Rechenregeln**

1. Assoziativgesetz: $a(bc) = (ab)c \qquad a + (b + c) = (a + b) + c$
2. Kommutativgesetz: $ab = ba \qquad a + b = b + a$
3. Distributivgesetz: $a(b + c) = ab + ac \qquad a + (bc) = (a + b)(a + c)$
4. DeMorgan'sche Regel: $\overline{ab} = \bar{a} + \bar{b} \qquad \overline{a + b} = \bar{a}\bar{b}$

wobei Regel 4 die Dualität formuliert.

Es können im ganzen 16 verschiedene Wertetabellen für $y = f(a, b)$ aufgestellt werden. Es können aber wie zur erwarten alle aus den drei Grundoprationen abgeleitet werden. Vier dieser **abgeleiteten Grundoperationen** haben einige Bedeutung:

NAND

a	b	y
0	0	1
0	1	1
1	0	1
1	1	0

NOR

a	b	y
0	0	1
0	1	0
1	0	0
1	1	0

EXOR

a	b	y
0	0	0
0	1	1
1	0	1
1	1	0

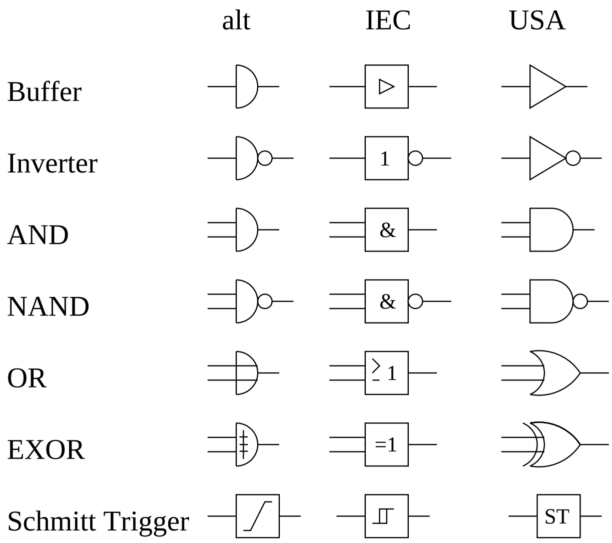
EXNOR

a	b	y
0	0	1
0	1	0
1	0	0
1	1	1

NAND und NOR sind die negierten von AND und OR. Sie haben darum eine spezielle Bedeutung, weil mit jedem von ihnen alle (!) Operationen aufgebaut werden können. Das exklusive OR, EXOR, heisst auch Antivalenz $y = (a \neq b)$ oder $y = a \oplus b$. Das invertierte davon heisst EXNOR oder Äquivalenz $y = (a = b)$ oder $y = \overline{a \oplus b}$. Es gilt im weiteren:

$$a \oplus b = a\bar{b} + \bar{a}b \qquad \overline{a \oplus b} = ab + \bar{a}\bar{b} \tag{6.3}$$

Die folgende Tabelle zeigt die wichtigsten Symbole der digitalen Schaltungstechnik. IEC ist die europäische Norm. Die mit veraltet bezeichnete Norm wird aber auch noch häufig verwendet.



Das Aufstellen der Gleichung aus der Wahrheitstabelle und die Entwicklung der Schaltung mit Hilfe der *diskjunktiven Normalform* soll an einem **Beispiel** demonstriert werden: Man stelle die Zahl der aktiven Bits von im ganzen 3 Bits a, b, c als eine Binärzahl y_0, y_1 dar (eine Aufgabe, die sich zum Beispiel bei einem Triggersystem der Teilchenphysik häufig stellt). Die Wahrheitstabelle lautet:

a	b	c	y_0	y_1
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
1	0	0	1	0
0	1	1	0	1
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

Die vollständige logische Beschreibung der Tabelle erfolgt nun in der disjunktive Normalform. Diese besteht aus einem grossen OR aller Zeilen, die eine 1 als Resultat haben. Jede Zeile wird dabei als Konjunktion (AND) aller Eingänge geschrieben, also Zeile zwei lautet zum Beispiel $y_0 = \bar{a}\bar{b}c$.

Die vollständige disjunktive Normalform (auch *sum of products* SOP genannt) unserer Tabelle lautet demnach:

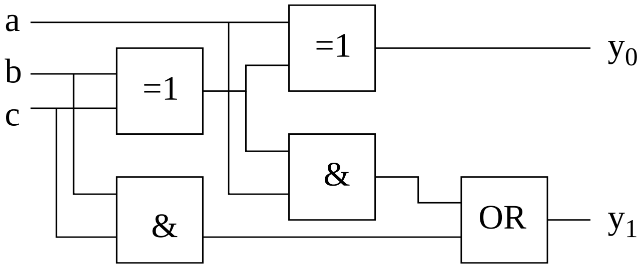
$$y_0 = \bar{a}\bar{b}c + \bar{a}b\bar{c} + a\bar{b}\bar{c} + abc \tag{6.4}$$

$$y_1 = \bar{a}bc + a\bar{b}c + ab\bar{c} + abc \tag{6.5}$$

Durch Anwendung der Rechenregeln erhält man daraus

$$y_0 = a \oplus (b \oplus c) \tag{6.6}$$

$$y_1 = a(b \oplus c) + bc \tag{6.7}$$



Wir brauchen also zwei EXOR, zwei AND und ein OR um dieses Schaltnetz zu realisieren, so wie es der nebenstehende Schaltplan zeigt. Es gibt natürlich mehrere Lösungen zu dem Problem, die Wahl hängt von vielen Details ab.

Alternativ dazu besteht die *konjunktive Normalform* (oder *product of sums* POS) aus der zur disjunktiven Form dualen Logik: Ein grosses AND wird aus den Zeilen mit den 0 als Resultat gebildet, wobei jede Zeile als OR aus den invertierten Werten geschrieben wird:

$$y_0 = (\bar{a} + \bar{b} + c) \cdot (\bar{a} + b + \bar{c}) \cdot (a + \bar{b} + \bar{c}) \cdot (a + b + c) \tag{6.8}$$
$$y_1 = (\bar{a} + b + c) \cdot (a + \bar{b} + c) \cdot (a + b + \bar{c}) \cdot (a + b + c) \tag{6.9}$$

Je nachdem ob die Resultatspalte mehr nullen oder mehr einsen enthält wählt man die konjunktive oder die disjunktive Normalform zur Beschreibung der Schaltung. (In unserem Beispiel sind es zufällig gleichviele!)

Diese Schaltungsentwicklung ("Synthese") kann natürlich mit Hilfe von Computerprogrammen ("hardware description languages") automatisiert werden (siehe Abschnitt 7.4). Im weiteren gibt es programmierbare Funktionsspeicher (PLD's) die direkt die disjunktive oder konjunktive Normalform programmierbar implementiert haben. 11.1.99

6.3 Innerer Aufbau von digitalen Bausteinen

6.3.1 Betriebsparameter und Auswahlkriterien, Bezeichnungen

Verschiedene Aufbauten der Grundschaltungen (Gatter) OR, AND und NOT aus konkreten elektronischen Bauteilen führen zu den verschiedenen Schaltungsfamilien. Für die Wahl der Familie sind unter anderen folgende technische Paramter relevant:

- Definition der Darstellung der Zustände (siehe 6.1) und ihr Störabstand.
- Die maximalen Eingangs- und die minimalen Ausgangsströme werden als *fanout* angegeben, das heisst die Zahl der Eingänge derselben Familie, die von einem einzigen Ausgang getrieben werden können.

- *Anstiegszeit* und *Abfallzeit* (*risetime*, *falltime*) der Ausgangssignale bei einer Änderung des logischen Ausgangszustandes. Dieser Wert ist stark von der Lastkapazität abhängig, er wird in der Regel bei einer Last von einer bestimmten Zahl von Eingängen von anderen Bausteinen derselben Familie angegeben. Beachte aber, dass die Lastkapazität auch von der Leitungslänge abhängt, lange Leitungen (auch auf p.c.b.'s) brauchen deshalb immer wieder buffers.

- *Verzögerungszeit*, *propagation delay*: die Zeit die es braucht von der Änderung des Eingangszustandes bis sich der Ausgangszustand ändert. Die propagation delay muss bei mehreren Ein- und Ausgängen im Prinzip für jede Kombination von Ein- und Ausgang separat angegeben werden.

Ausserdem müssen organisatorische Aspekte (Verfügbarkeit, Service- und Austauschmöglichkeiten, Preis usw.) sowie weitere technische Eigenschaften (Zuverlässigkeit, Betriebsspannungen, Stromverbrauch, Montagetechnik usw.) berücksichtigt werden.

Logikbausteine werden mit dem Code FF-74TTTnnnPPP bezeichnet. Dabei bedeutet

- FF das Herstellerkürzel

- TTT die Schaltungsfamilie

- nnn eine Zahl, die die eingebaute Logik definiert

- PPP Gehäuse und Betriebstemperaturbereich (military: $-25\ldots+125°$ C, industrial: $-25\ldots+85°$ C, commercial: $0\ldots+70°$ C)

Beispiel: SN74ALS00N: Hersteller Texas Instruments, Familie advanced low power schottki, 00: vier unabhängige NAND gates, Plastikgehäuse im DIL-14 Format.

Die Details der Kodierung sind herstellerabhängig, eine etwas ausführlichere aber auch unvollständige Beschreibung findet sich in [Nühr98] Seite 3516. Eine ziemlich vollständige Liste der Hersteller Prefixes FF findet man in [Horo97], Seite 1069. Hat man damit den Hersteller identifiziert, schaut man am besten in den Unterlagen des entsprechenden Herstellers (auf dem internet) nach, wie dieser genau die Kodierung macht. Glücklicherweise sind wenigstens die Logik Nummern nnn einigermassen einheitlich verwendet.

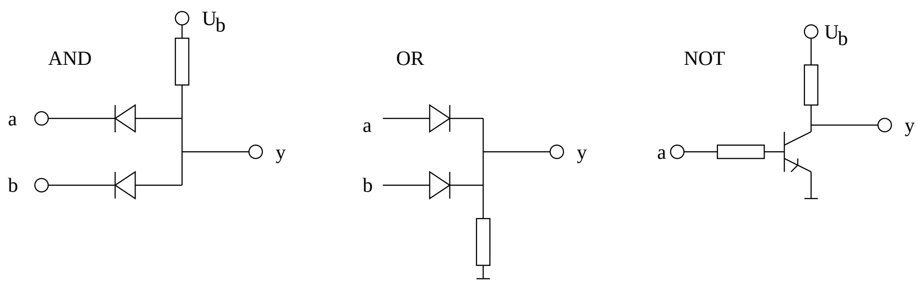
6.3.2 Relais

Ein *Relais* besteht aus einem Eisenkern, der von einem oder mehreren Spulen (logischen Eingängen) magnetisiert wird. Das Magnetfeld betätigt dann einen oder mehrere Schalter, die damit ihren Zustand von offen auf geschlossen (oder umgekehrt: "Auskontakt") ändern. AND und OR Schaltungen bestehen aus zwei Relais deren Schaltkontakte in Serie oder parallel geschaltet werden. NOT Schaltungen bestehen aus einem Relais mit einem Auskontakt.

Relaisschaltungen sind langsam, störanfällig und haben einen immensen Strom- und Platzbedarf. Ihr entscheidender Vorteil besteht darin, dass Eingangs- und Ausgangsstromkreis galvanisch völlig getrennt sind, sie werden deshalb als Signalübertrager auf hochgespannte oder variable Potentiale und manchmal für das Schalten grosser Ströme immer noch verwendet.

6.3.3 Dioden und Transistoren

Mit Hilfe nur von Dioden und Widerständen können AND und OR Schaltungen realisiert werden, für NOT braucht es mindestens einen Transistor.

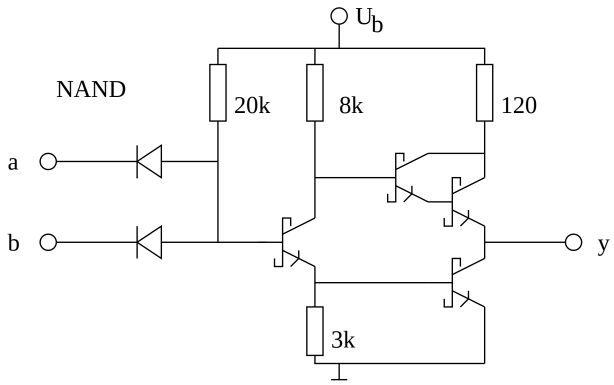


Diese Schaltungen werden vorallem dazu verwendet, an bestehenden Systemen auf einfache Art noch zusätzliche einzelne Verknüpfungen einzuführen. Sie sind allerdings etwas langsam (Anstiegszeit RC, R aus der Schaltung, C die Lastkapazität) oder bei kleindimensioniertem R ziehen sie von den Ausgängen eine grosse Last.

Es können im Prinzip beliebig viele Dioden am Eingang der UND Schaltung verwendet werden. Man spricht in diesem Zusammenhang vom *wired OR* (*eine* aktive Diode genügt, um den Ausgang auf Low zu bringen: invertierte Logik). Dieses Prinzip eignet sich besonders, um viele Eingänge, die an gegraphisch verteilten Orten sitzen, miteinander zu verknüpfen.

6.3.4 TTL

Die *TTL* (Transistor – Transistor – Logik) Technik besteht aus Grundschlatungen von mehreren Transistoren. Es gibt verschiedene Unterfamilien, die sich durch die Details der Dimensionierung der Art- und Schaltung der verwendeten Transistoren unterscheiden.



Das Bild zeigt die Grundschaltung eines *LS-TTL* NAND – Gatters 7400. LS steht für Low Power (relativ hochohmig dimensionierte Widerstände) und Schottki. Letzteres bezieht sich auf einen Schottkikontakt (Halbleiter Metallübergang) zwischen Basis und Kollektor der verwendeten Transistoren (S-förmiges Symbol). Damit wird verhindert, dass die Transistoren durch negative BC Spannung in die Sättigung gehen, die Geschwindigkeit der Schaltungen wird damit stark erhöht.

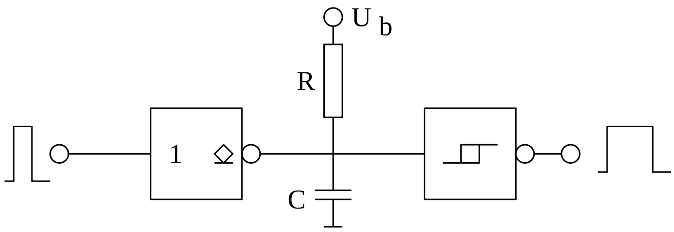
Typische Verlustleistung eines LS Gatters beträgt 2 mW, propagation delay 15 ns. In den Eingang fliessen im H Zustand max. 20 μA, aus dem L Zustand max. 400 μA. Die Ausgänge können mindestens 8 mA im L und 400 μA im H Zustand treiben. Die ergibt ein garantiertes fanout von 20. Dank den vergleichbaren Eigenschaften von Leuchtdioden ergibt sich in der Praxis, dass man mit TTL Ausgängen direkt LED's (bevorzugt grüne: $U_f \approx 2.2V$) gegen Masse ohne zusätzliche Vorwiderstände betreiben kann, was oft sehr hilfreich ist.

Die ALS (advanced Low Power Schottki) – Typen erlauben durch zusätzliche Schaltungstricks eine propagation delay von 5 ns, ohne den Stromverbrauch zu erhöhen. AS (advanced Schottki) haben niedriger dimensionierte Widerstände und sind deshalb schneller: propagation delay 1 bis 2 ns, Stromverbrauch 10 mW.

Beachte aber, dass je schneller die Schaltkreise werden, umso höher ist die Gefahr des Übersprechens (X-talk). Auch können die im Schaltzeitpunkt erhöhten Ströme wegen der Zuleitungsinduktivität die Spannungsversorgung kurzzeitig reduzieren. Um diesen Effekt klein zu halten, muss möglichst nahe an jedem digitalen Baustein ein Stützkondensator zwischen Masse und Betriebsspannung geschaltet werden: 10 bis 100 nF Keramik pro Chip. Trotzdem begrenzen diese Art von Problemen den Einsatz von TTL bei sehr grossen Schaltfrequenzen.

6.3.5 Open Collector und Tristate

Um die Verknüpfung sehr vieler Signale in einem grossen OR oder AND mittels "wired OR" zu vereinfachen, werden viele der Schaltkreise auch mit *open collector* angeboten (Abkürzung: OC, Symbol siehe im Beispiel unten). Dabei werden am standard Schaltkreis einfach die beiden oberen Darlington Transistoren weggelassen. Zum Beispiel hat das vierfach NAND mit open collector die Nummer 7401. Damit erkauft man sich allerdings den Nachteil, dass wie in der Diodenschaltung oben beschrieben, die Anstiegszeit des Signales relativ langsam wird.



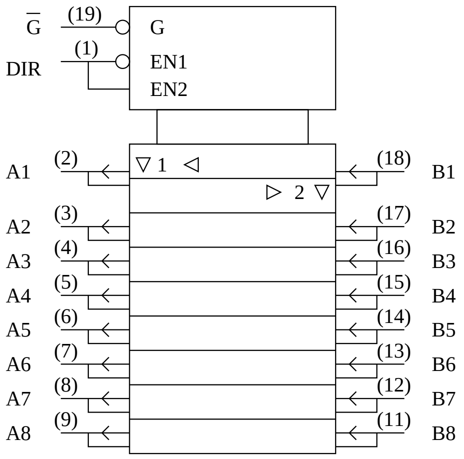
Anwendung: Signalverzögerung mittels eines open Collectors und eines Schmitt – Triggers. Die abfallende Flanke wird um etwa $0.5 \cdot RC$ verzögert.

Eine weitere Anwendung der open collector Schaltung sind interfaces zu Systemen mit anderen Spannungsversorgungen: Der Pull–up Widerstand wird einfach an die neue Spannung gehängt, und schon ist das Interface fertig. Beachte aber V_{CE} max.

Bei der *tristate* Ausgangsschaltung werden die oberen und der untere Transistor in der Ausgangsstufe logisch separat angesteuert, sodass neben den bereits bekannten Ausgangszuständen L und H es auch möglich ist, mittels eines separaten Eingangs "Enable" alle Transistoren zu sperren, und somit einen hochohmigen Zustand Z zu erreichen.

Mit Tristate Ausgängen lässt sich somit ein echter *Busbetrieb* realisieren, das heisst die gleiche Signalleitung kann von wahlweise von verschiedenen Gattern betrieben werden. Damit wird es ermöglicht, mehrere Systeme am gleichen Kabel zu betreiben, vorausgesetzt die Enable Leitungen der verschiedenen Schlatkreise werden auf koordinierte Art betrieben. 13.1.99

Ein spezifische Anwendung der Tristate Technik sind die sogenannten *Bustransceivers.* Sie dienen als Leitungsverstärker sowie als "Tor" eines Gerätes zu einer Busleitung, falls Daten in beide Richtungen übertragen werden sollen.

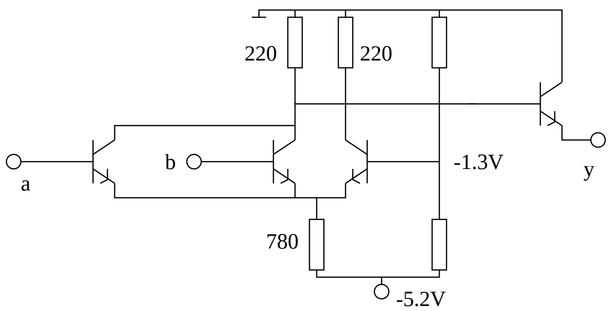


Die nebenstehende Abbildung zeigt einen 8-fach nichinvertierenden Bustranceiver (Typ '645). Das auf der Spitze stehende Dreieck steht für den Tristate Ausgang, das liegende Dreieck für den Verstärker, da der '645 einen relativ hohen Ausgangsstrom liefern (typisch 24 mA) und damit ein fanout von 60 treiben kann, sodass er lange Leitungen mit hohen Kapazitäten bzw. vielen Anschlüssen bedienen kann. Der DIR Eingang bestimmt die Richtung der Übertragung, mit $\bar{G}$ lassen sich alle Ein- und Ausgänge ausschalten.

Die Zahlen in Klammern geben die Anschlussnummer (pin) am Gehäuse an (in der Regel im Gegenuhrzeigersinn).

6.3.6 ECL

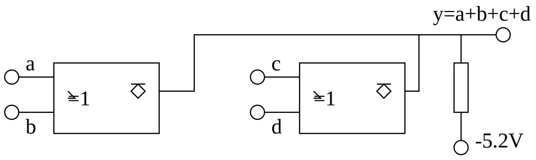
Die ECL Technik wurde speziell mit dem Ziel geschaffen, den Stromverbrauch unabhänig vom Schaltzustand zu halten, und damit die durch Schaltströme ausgelösten "spikes" auf der Strom- und Massezuführung zu vermeiden.



Das nebenstehende Bild zeigt die Grundschaltung eines NOR ($y = \overline{a+b}$) in *emittergekoppelte Logik* (*ECL*) der 10k Serie. Es handelt sich um einen Differenzverstärker, dessen Emitter- und Kollektorwiderstände so dimensioniert sind, dass die Transistoren nicht in Sättigung gehen können.

Der Ausgang kann direkt an weitere Gatter angeschlossen werden. Bleibt er offen, muss er allerdings mit 520 Ω nach -5.2V abgeschlossen werden, sonst sieht man kein Signal!

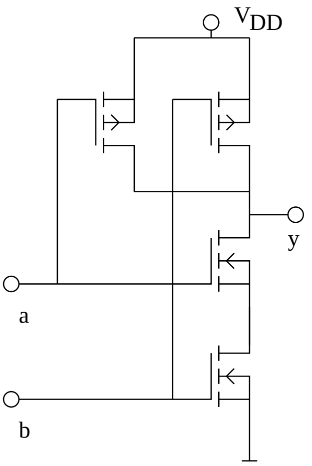
ECL Schaltkreise sind sehr schnell, man kann unter 1 ns propagation delay (ECL 100k) erreichen, was mit einem Stromverbrauch von 40 mW pro gate erkauft wird. Grössere ECL Schaltungen brauchen normalerweise Wasserkühlung. Da die Summe der Stromflüsse im ECL gate im wesentlichen unabhängig von dem Schaltzustand ist, erlaubt ECL einen sehr störsicheren Betrieb auch bei ganz hohen Schaltfrequenzen. Die kleinen Spannungsamplituden erlauben eine schnelleres Umladen der Streukapazitäten, was ebenfalls zu einer Geschwindigkeitsverbesserung führt. Die neueste Generation der ECL Familie ist die ECLinPS–Lite (PS steht für picosekunden), die propagation delay von 220 ps und Schaltfrequenzen bis zu 2.8 GHz erlaubt. Das erkauft man mit einer Verlusteistung von 73 mW und Kosten von 32$ pro gate.



Der offenen Emitterausgang ermöglicht es, leicht "wired OR" Verknüpfungen zu machen. Beachte das Symbol für den offenen Emitterausgang.

In der Kern- und Teilchenphysik wird der sogenannte *NIM - Standard* verwendet. Dabei handelt es sich im wesentlichen um ECL Technik, die Ausgänge für Kabelanschlüsse sind dabei mit einem zusätzlichen Ausgangstreiber versehen, sodass sie in der Lage sind 50 Ω Kabel gegen Masse zu treiben. Der L Zustand wird 0 V, der H Zustand -0.8 V.

6.3.7 CMOS



Complementary MOS Technik (*CMOS*) stellt heute den mit grossem Abstand überwiegendsten Teil aller digitaler Elektronik. Das Bild zeigt ein standard NAND gate mit 4 enhancement (selbstsperrenden) MOSFETs. Im Ruhezustand fliesst paraktisch kein Strom ($P < 10\mu$W). Dafür sind allerdings die Umladungsströme im Schaltmoment relativ gross, sodass sich eine im wesentlichen lineare Frequenzabhängigkeit der Verlustleistung ergibt. Typischerweise bei einigen MHz wird der Stromverbrauch grösser als der von LS–TTL. Die Versorgungsspannung V_{DD} liegt im Bereich 2...5.5 V (VHC), 3...18 V (4000B Serie) oder 4.5...5.5 V (HCT).

Die progagation delays sind sehr unterschiedlich, die schnellsten (AC) erreichen unter 3 ns, haben aber eine sehr schlechte X-talk Charakteristik innerhalb des gleichen chips (substrate bouncing), sodass sie nur inspeziellen Anwendungen verwendet werden sollten. HC haben typisch 10 ns und funktionieren sehr stabil, sodass sie für die meisten Anwendungen die gute Wahl sind. Die Version HCT ist wie HC, hat aber TTL – kompatible Eingangscharakteristik. Die neueste Gerneration dieser Familie ist LCX, die bei 3 V Speisung nur noch halb so viel Power zieht (0.3 mW bei 1 MHz) wie HC und trotzdem bis zu Frequenzen von 200 MHz läuft (3.5ns propagation delay). Besonders angenehm ist, dass die Ein- und Ausgänge trotz 3V Speisung mit 5V betrieben werden können. LCX wurde von den Firmen Motorola, Toshiba und National gemeinsam spezifiziert.

Für sehr hochintegrierte Schaltungen wird die NMOS Technik verwendet, die nur aus NMOS FETs besteht, und noch einfachere Schaltungen und damit höhere Logikdichte erlaubt.

Eine wesentliche Begrenzung der Schaltgeschwindigkeiten stellen die umzuladenden Streukapazitäten dar. Da Leitungen innerhalb eines chips um Grössenordnungen kürzer sind, als auf einem p.c.b., können innerhalb von chips noch wesentlich höhere Geschwindigkeiten (viele hundert MHz) erreicht werden. Darauf gründet unter anderem der rasante Fortschritt in der Geschwindigkeit von Mikroprozessoren (80X86).

Die open collector Versionen (sollten eigentlich open drain heissen) gibt es teilweise ebenfalls, zum Beispiel der MC74HC03, ein 4–fach 2 input NAND mit open collector Ausgängen von Motorola. Ebenso die Tristate Versionen, der CD74HC640 ist ein tristate 8–facher invertierender Bustransceiver von Harris/RCA.

Einige Grundreglen bei der Verwendung von CMOS Bauteilen müssen beachtet werden:

1. Es müssen alle unbenützten Eingänge auf definiertes Potential angeschlossen werden, auch solche von nicht benützten Logikteilen, da diese bereits mit minimaler Ladung am Eingang umschalten können.

2. Alle Eingänge haben Schutzdioden auf Masse und V_{DD} und sind deshalb nicht sehr empfindlich auf statische Elektrizität. Führt man aber zu grosse Spannungen an die Eingänge (statisch oder Überschwingen), können diese Dioden durchbrennen. Es kann sich auch ein sogenanntes *Latchup* ergeben (Der NFET und der PFET am Eingang werden gleichzeitig leitend $\Rightarrow$ Kruzschluss). Um das zu verhindern, sollten bei Eingängen, die über Stecker oder von längeren Leitungen kommen, immer 10kΩ Widerstände in Serie geschaltet werden. Bei den HC Typen ist ein Seriewiderstand von 120 Ohm eingebaut, der die Latchup Wahrscheinlichkeit stark reduziert.

3. Die erwähnten Schutzdioden können dazu führen, dass eine logische Schaltung scheinbar funktioniert, ohne dass eine Speisspannung an den IC's angeschlossen ist, da Strom von den logischen Eingängen über die Schutzdioden zu den Speisanschlüssen fliesst. Das ist aber natürlich kein zuverlässiger Betrieb!

4. CMOS ziehen viel Strom beim Schalten, man muss also hier auch wie bei TTL typisch 100 nF Keramik pro Chip vorsehen, um die Speisung zu stabilisieren.

5. Die Signalqualität wird umso schlechter je länger die Leitung und je mehr Eingänge an einen Ausgang angeschlossen werden. Für schnelle Schaltungen muss deshlab die optimale Zahl von zusätzlichen Signaltreibern (die natürlich selbst eine gewisse propagation delay haben) bestimmt werden.

6. Bei komplexeren Aufbauten wird die Spannungs- und Masseführung zum Killer des Systems. Ohne je eine durchgehenden Masse- und V_{DD} Schicht im multilayer p.c.b. zu spendieren besteht keine Chance für einen zuverlässigen Betrieb!

6.3.8 Vergleich der Logikfamilien und –Generationen

(siehe http://mot-sps.com/logic/which_is_best.html)

Motorola Logic Families, Which Is Best for You?

By Gary Tharalson, Motorola, Inc., Mesa, AZ

Typical Commercial Parameter (0°C to +70°C)	Logic Families												
	TTL				CMOS					ECL			
	LS	ALS	ABT	FAST	MG	HC	FACT	LVC	LCX	10H	100K	ECL in PS(3)	E-Lite
Speed													
Gate Prop Delay (ns)	9	7	2.7	3	65	8	5	3.3	3.5	1	0.75	0.33	0.22
Flip-Flop Toggle Rate (MHz)	33	45	200	125	4	45	160	200	200	330	400	1,000	2800
Output Edge Rate (ns)	6	3	3	2	50	4	2	3.7	3.6	1	0.7	0.5	0.25
Power Consumption Per Gate (mW)													
Quiescent	5	1.2	0.005	12.5	0.0006	0.003	0.0001	0.003	1E-04	25	50	25	73
Operating (1 MHz)	5	1.2	1.0	12.5	0.04	0.6	0.6	0.8	0.3	25	50	25	73
Supply Voltage (V)	+4.5 to +5.5	+4.5 to +5.5	+4.5 to +5.5	+4.5 to +5.5	+3 to +18	+2 to +6	+1.2 to +3.6	+2 to +3.6	+2 to +6	-4.5 to -5.5	-4.2 to -4.8	-4.2 to -5.5	-4.2 to -5.5
Output Drive (mA)	8	8	32/64	20	1	4	24	24	24	50 ohm load	50 ohm load	50 ohm load	50 ohm load
5V Tolerant													
Inputs	N/A	N/A	N/A	N/A	N/A	N/A	N/A	Yes	Yes	N/A	N/A	N/A	N/A
Outputs	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No	Yes	N/A	N/A	N/A	N/A
DC Noise Margin (1)													
High Input %	22	22	22	22	30	30	30	30	30	27	41	28/41	33
Low Input %	10	10	10	10	30	30	30	30	30	31	30	31/31	33
Packaging(4)													
DIP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	No
SO	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	Yes
LCC	No	Yes	No	Yes	No	No	Yes	No	No	Yes	No	Yes	No
SSOP	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No
TSSOP	No	No	No	No	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Functional devices Types	190	210	50	110	125	103	80	35	27(2)	64	44	48	40
Relative 1-25 Gate Price U.S. $	.90	1.00	1.60	1.00	.90	.90	1.50	1.80	1.80	2.00	10.00	28.00	32.00

18.1.99

(1)Typical noise margin expressed as a percentage of typical output voltage swing.
(2)Announced plans for Motorola offering.
(3)ECLinPS is available in both 10H and 100K compatible versions.
(4)A "Yes" may not include all devices within a family.

Manufacturers Data Books Referenced:
Motorola, DL121, DL122, DL129, DL131, DL138, DL140, BR1339
Texas Instruments, SDAD001B
Phillips Semiconductor, IC23, IC24
National Semiconductor, F100K

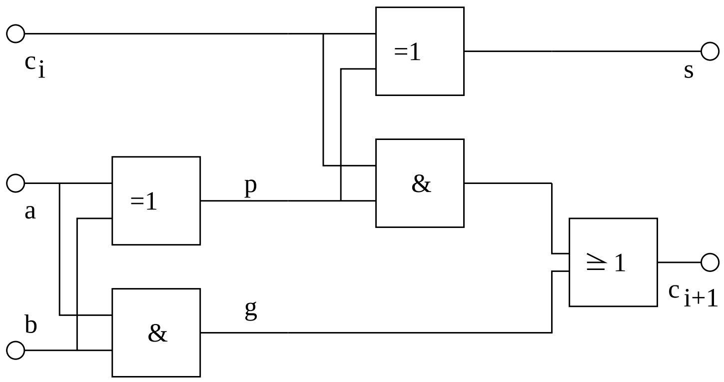
6.4 Beispiele von digitalen Grundschaltungen

6.4.1 Halbaddierer und Volladdierer

Führt man sich die Wahrheitstafel der Addition von 2 binären Stellen zweier Zahlen a und b vor Augen, sieht man

$$s = a \oplus b \qquad c = ab \tag{6.10}$$

wobei s die Summe und c das carry (Übertrag) darstellt. Eine Schaltung, die diese Logik enthält heisst ein *Halbaddierer*, da sie nicht in der Lage ist, allfällige carries von der vorhergehenden Stelle zu übernehmen.



Für einen *Volladdierer* (*Fulladder*) lauten die Gleichungen

$$\begin{aligned} s &= a \oplus b \oplus c_i \\ c_{i+1} &= ab + c_i(a \oplus b) \end{aligned}$$

(c_i das carry aus der vorhergehenden Addition) und braucht demnach schon eine etwas kompliziertere Logik.

Das Zwischenresultat $g := ab$ heisst *generate*, $p := a \oplus b$ heisst *propagate*.

Um zwei mehrstellige Dualzahlen addieren zu können, braucht man für jede Stelle einen Volladdierer. Der Aufwand an gates geht proportional zu den Anzahl Stellen, ist also kein Problem. Hingegen geht die gesamte Verzögerungszeit wegen den durchzuschlaufenden Carries ebenfalls mit den Anzahl Stellen linear hoch, man spricht vom seriellem Übertrag oder von *ripple carry*.

Die Situation lässt sich verbessern, indem man in die Gleichung für das carry $c_{i+1} = g_i + p_i c_i$ alle vorhergehenden Zwischenresultate einsetzt, und auf die disjunktive Normalform bringt, man erhält dann Ausdrücke der Form

$$c_{i+1} = \sum_{k=1}^{i} (\prod_{l=k}^{i} p_l) \cdot g_{k-1} \tag{6.11}$$

Diese Summe von Produkten lässt sich also immer mit zwei gate propagation delays darstellen. Dazu kommt noch ein je ein delay von der Generation der p_i und g_i, sowie die EXOR am Ausgang für die Resultatbildung, sodass eine propagation delay von 4 gates unabhängig von der Anzahl Stellen entsteht. Man spricht vom *carry look ahead*.

Ein typischer 4 bit full adder mit carry look ahead ist der CD74HC283, der es auf eine Addierzeit von ca. 30 ns bringt.

In der Teilchenphysik hat man oft das Problem, dass man für den Trigger eines Experimentes (schnelle Entscheidung, ob ein Ereignis gut oder schlecht ist) sehr viele einzelne Messkanäle (z.B. von einem Calorimeter) addieren muss. Die offensichtliche Lösung ist eine Baumstruktur aus einzelnen Addierer zu bauen. Für die korrekte Addierung aller möglichen Werte muss man bei jeder Stufe des Baumes eine Stelle mehr vorsehen (!) Will man zum Beispiel 512 Acht–Bit Zahlen addieren, so braucht man neun Addierstufen, die erste noch für 8 bit, die letzte für 17 bit. Da jedoch in diesen Experimenten meistens nur wenige Kanäle von null verschiedene Werte haben, lässt man das carry oft weg, implementiert nur 8 bit adders und anstelle des carry ein overflow – Signal, das ein einfaches OR aller carries ist.

Die *Subtraktion* zweier Zahlen lässt sich gleich wie eine Addition im obigen Sinne behandeln, wenn man für negative Zahle die *Zweierkomplement* (*Two's complement*) – Darstellung wählt. Das Zweierkomplement einer Dualzahl erhält man durch Inversion aller Stellen und anschliessender Addition von 1. In einer Achtbit Darstellung wird -1 = 11111111 und -16 = 11110000. Bei höherer Stellenzahl der Darstellung müssen also bei negativen Zahlen weitere einsen vorangestellt werden (*sign extension*). Das ist oft der Fall beim Einlesen von Resultaten eines ADC, da letztere oft wesentlich weniger Bits haben als der verarbeitende Computer!

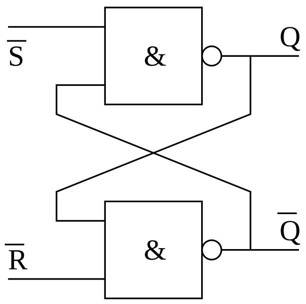
Für floating point Zahlen werden Mantisse und Exponent seperat dargestellt und bearbeitet. IEEE hat die Details dieser Darstellung normiert. Dazu sei auf das Computer–hardware Praktikum von Prof. V. Lindenstruth verwiesen.

6.4.2 Flip–Flops

Alle bisherigen logischen Schaltungen waren *Schaltnetze* (*combinatorial logics*), das heisst ihr Ausgangszustand hängt nur von den am Eingang anliegenden Signalen ab, sofern man die maximale propagation delay abgewartet hat.

Von einem *Schaltwerk* (*sequential logics*) spricht man, wenn die Ausgangszustände ausserdem auch von der Vorgeschichte abhängen, man braucht also Speicher. Stellt man solche Speicher aus logischen Gattern her, spricht man vom *Flip-Flop*. (ALternativ kann man auch andere physikalische Effekte heranziehen, zum Beispiel geladene Kondensatoren $\Rightarrow$ dynamisches RAM).

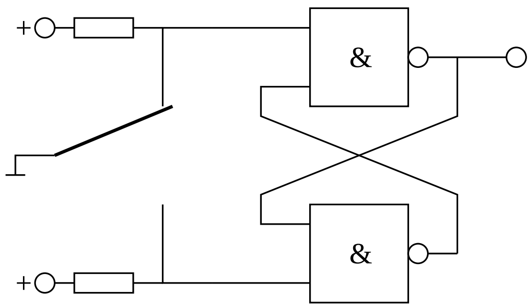
RS Flip–Flop



Den einfachsten Speicher erhält man durch nebenstehende Schaltung. Die dazugehörige Wahrheitstabelle lautet

S	R	Q	$\overline{Q}$
0	0	Q_{-1}	$\overline{Q}_{-1}$
1	0	1	0
0	1	0	1
1	1	1	1

Die erste Zeile entspricht dem Speicherzustand. Der gespeicherte Wert hängt davon ab, ob unmittelbar vor dem Eintreffen des Speicherzustandes die Werte von Zeile 2 oder 3 vorhanden waren. Zeile 4 führt nicht zu einem eindeutigen Speicherzustand.

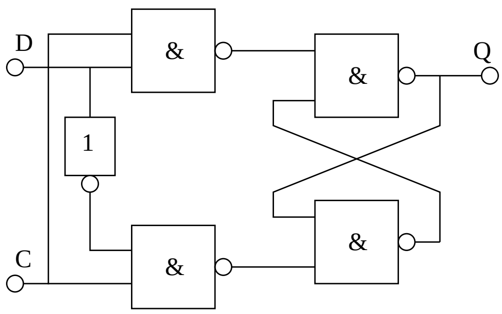


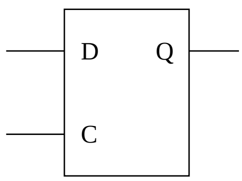
Anwendung: Entprellen eines Schalter. Beachte die Widerstände!

20.1.99

Transparentes D–Latch

Die viel häufigere Anforderung an einen Speicherbaustein besteht darin, den Zustand eine Leitung D zu einem Zeitpunkt einzufrieren, der durch die Flanke einer Taktleitung C (*Clock*) bestimmt ist.

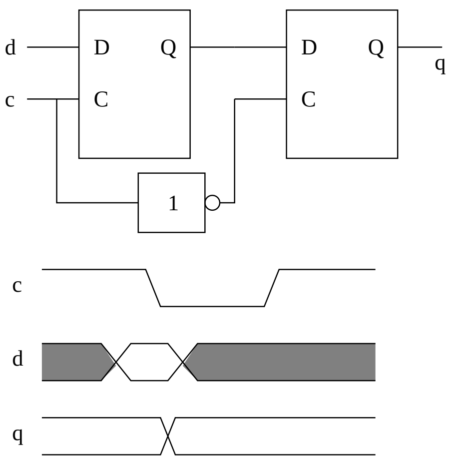




Die Schaltung heisst transparentes Latch (manchmal auch transparentes D–Flipflop oder einfach D–Latch genannt). Solange das Clock–Signal anliegt widerspiegelt der Ausgang den Eingang. Geht das Clock–Signal auf L, stellt der Ausgang den unmittelbar davor bestehenden Zustand von D dar. Das rechte Bild zeigt das Symbol für ein D–Latch.

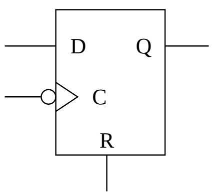
Bei diesen Schaltungen sind die früher besprochenen Setup- und Holdzeiten natürlich wichtig (siehe Abschnitt 6.1). Beispiel Datenblatt MC74VHC573.

Das Master–Slave Prinzip



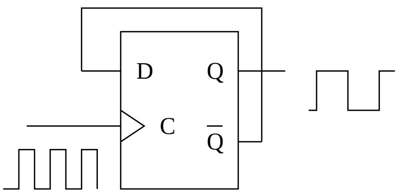
Setzt man zwei D–Latches in Serie hintereinander, wobei das zweite Latch die invertierte clock bekommt, erhält man ein *master–slave* Flipflop. Liegt die clock auf H, ist das hintere Latch transparent, das vordere speichert. Wechselt die clock auf L, friert das hintere Latch ein und das vordere wird transparent und zeigt den gespeicherten Zustand. Geht die clock wieder auf H, speichert das vordere Latch den alten Zustand, das hintere wird nun transparent.

Der Clock–Übergang von L auf H bleibt also ohne Auswirkung auf den Ausgangszustand, während beim Übergang von H auf L die Eingangsdaten gespeichert, und am Ausgang dargestellt werden.



Das nebenstehende Bild zeigt das Symbol für ein Master – Slave D–Fliflop, oft auch *edge triggered* D-Flipflop oder einfach (etwas verwirrend) nur D-Flipflop genannt. Mit dem Dreieck wird die Flankensensitivität ausgedrückt. Oft ist ein weiterer Reseteingang R vorhanden, der den Speicherzustand löscht, was zum Beipsiel bei der Initialisierung relevant ist.

Toggle



Das nebenstehende Bild zeigt ein *toggle*, das aus einem edge–triggered D–Flipflop aufgebaut ist. Bei jeder positiven Clockflanke ändert sich der Zustand des Eingangs in sein Invertiertes. Man spricht deshalb auch von einer *divide–by–2* Schaltung, oder von einem einstelligen Zähler.

Steuerbare Toggles haben noch einen Steuereingang T, der Ausgang wird nur umgeschaltet, wenn $T = H$ ist.

6.4.3 Zähler

Einen *asynchroner Zähler* (*ripple counter*) erhält man durch einfache Serieschaltung von ungesteuerten Toggles und zwar soviele, wie bits zu zählen sind. Ist die resultierende Zahl auszuwerten, besteht in dieser Schaltung der Nachteil, dass die Ausgangswerte nicht gleichzeitig ankommen, sondern bei jeder Stelle die Laufzeit t_T durch ein Toggle dazu zu zählen ist. Das letzte Bit wird erst mit einer Verzögerung von $n \cdot t_T$ seinen korrekten Wert annehmen. Für divide–by–2^n Schaltungen ist das allerdings irrelevant, und da kommt diese Methode vor allem für sehr viele bits zum Zug.

Beispiel: Der CD74HC4020 ist ein 14 stufiger asynchroner Binärzähler mit gemeinsamem Reset und läuft bis etwa 50 MHz. Die Propagation delay von einer Stufe beträgt aber etwa 6 ns, bei hohen Frequenzen zeigen die einzelnen Stellen also gar nie einen gemeinsamen korrekten Wert.

Legt man Wert auf korrekte, parallel zu verarbeitende Resultate, dann muss man einen *synchronen Zähler* wählen. Dabei werden gesteuerte Toggles verwendet, deren Takteingänge alle miteinander verbunden sind. Die Bedingung für das Zählen einer bestimmten Zelle wird durch ein Schaltnetz aus den Q_i gebildet und auf die entsprechenden Steuereingänge gegeben. Bei sehr vielen Stellen werden diese Netze allerdings kompliziert. Der CD74HC161 ist ein 4-Bit Binärzähler mit Reset Eingang, der mit einer beliebigen Zahl vorgeladen werden kann (*presettable Counter*). Er liefert auch einen Carry out und besitzt zwei enable Eingänge, sodass mehrere '161 kaskadiert werden können.

4-bit *BCD-Zähler* (*binary coded decimal*, auch *8421 Code*) zählen nur bis 9 (statt $2^4 - 1 = 15$) und beginnen dann wieder bei 0, entsprechen also einer Dezimalstelle. Sie sind für direkte Anzeigen geeignet. Beispiel HC160.

Die Serie HC190 bis HC194 hat zusätzlich einen Steuereingang für Vorwärts– Rückwärtszählung (*up/down counter*).

6.4.4 Schieberegister

Setzt man mehrere edge–triggered D-Flipflop in Serie (je Q mit D verbinden, aber für alle eine gemeinsame clock) erhält man ein *Schieberegister*. Je nach Anschlusssituation spricht man von einem serial in – parallel out (Beispiel HC164) oder parallel in – serial out (HC165).

Schieberegister kommen insbesondere bei der Umwandlung von serieller Datenübertragung in parallele oder umgekehrt zum Einsatz, aber auch für Pipelinearchitekturen, wo Daten für eine bestimmte Zahl von Taktfrequenzen zwischengespeichert werden müssen.

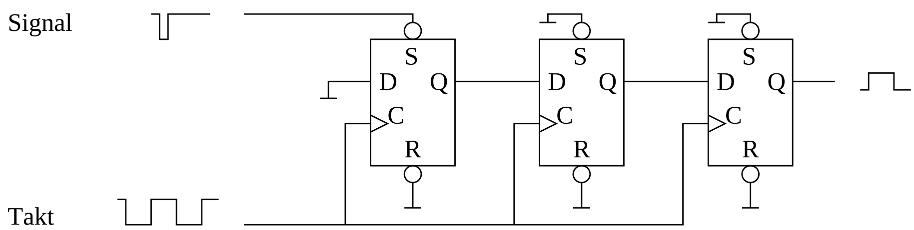
6.4.5 Synchronisation

Ein für den Expermentalphysiker wichtiger Aspekt der digitalen Elektronik ist die Ereignissynchronisation: Digitale Systeme wie Computer funktionieren mit einer festen internen Taktfrequenz. Das physikalische Signal, das von einem Detektor kommt, weiss von diesem Takt in der Regel nichts, das heisst es kommt im Vergleich zum Takt zu einer beliebigen, nicht vorhersehbaren Phasendifferenz.

Im Prinzip löst man das Problem mit einem D-Flipflop, an dessen Ausgang eine Signal entseht, dass synchron zur Taktfrequenz ändert. Es ergeben sich damit aber zwei Probleme:

1. Detektorpulse, die kürzer als eine Clockphase sind, gehen unter Umständen verloren.
2. Wenn die Setup- oder Hold – Zeiten verletzt werden, dann kann es unter Umständen zu einem fehlerhaften Zustand des flipflops kommen (metastabilen Zustand), bei dem der Ausgang keinen eindeutigen logischen Level mehr annimmt. Dieser Zustand kann einige Zeit andauern (μsec, msec...).

Das folgende Bild zeigt eine Lösung für die beiden Probleme. Als D-Flipflop wird zum Beispiel der HC74 verwendet, die $\bar{R}$ und $\bar{S}$ – Eingänge löschen, bzw. setzen das Flipflop, und zwar unabhängig vom Zustand der D und C Eingänge.



25.1.99

synchroner Aenderungsdetektor, TS page 260,

synchroner Taktein- und ausschalter, TS page 260

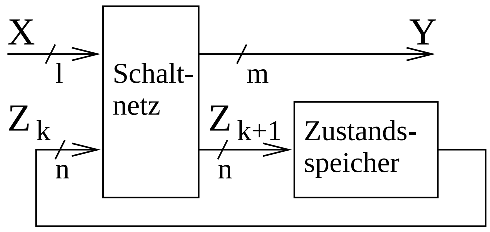
Kapitel 7

Höhere digitale Systeme

Wie bereits diskutiert unterscheiden wir zwischen *Schaltnetzen* (*combinatorial logic*) und *Schaltwerken* (*sequential logic*). Während Schaltnetze nur vom momentanen Eingangszustand abhängen (abgesehen von allfälligen gate propagation delays), erinneren sich Schaltwerke an die Vergangenheit, das heisst sie brauchen einen Speicher. Ändert sich der Ausgang eines Schaltwerkes nut bei festen Taktzeitpunkten spricht man von einem *synchronen Schaltwerk*.

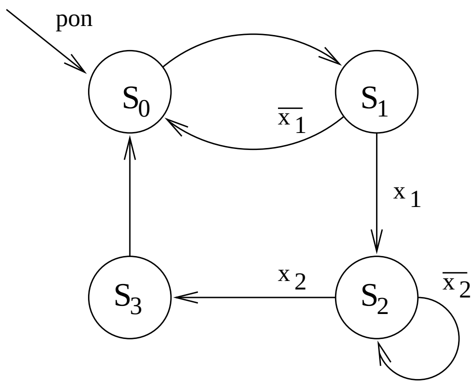
7.1 Finite State Machines

Finite State Machines (FSM) oder *endliche Automaten* sind synchrone Schaltwerke, die nur eine endliche Zahl von verschiednen Zuständen einnehmen können. Sie lassen sich schematisch wie folgt darstellen:



X, Y, und Z_k werden Eingangsvektor, Ausgangsvektor und Zustandsvektor (oder kurz Zustand) genannt, der letztere stellt also die Information dar, die für jeden Taktzeitpunkt gespeichert werden.

Eine FSM mit einem n dimensionalen Zustandsraum kann also 2^n verschiedene Zustände einnehmen. Welcher Zustand Z_{k+1} auf den Zustand Z_k folgt, hängt auch vom Eingangsvektor X (auch *Qualifier* genannt) ab. (Beachte, dass die bereits behandelten Zähler als FSM ohne Eingangsvariablen betrachtet werden können).



Oft verwendet man zur Beschreibung der Zustandsfolge das *Zustandsdiagramm*. Die Kreise bedeuten die einzelnen Zustände, die als Pfeile dargestellten Übergänge hängen jeweils vom Zustand einer Eingangsvariable ab, diese Übergangsbedingungen werden neben dem Pfeil notiert. Unbezeichnete Pfeile sind unbedingte Übergänge, die jeweils direkt im nächsten Takt stattfinden.

Ist keine der Bedingungen erfüllt, bleibt das System in seinem Zustand (Beispiel S_2). Diese Tatsache wird manchmal noch mit einer Schlaufe mit der Wartebedingung verdeutlicht. Jede FSM braucht eine Intialisierung, die zum Beispiel mit *pon* (Power on) bezeichnet wird.

Als alternative Darstellung findet das *Flussdiagramm* analog wie in der Informatik Verwendung. Es eignet sich aber eher für eine Implementation mit einem sequentiellen Programm.

Anwendungsbeispiel: Verkehrsampel

Realisierung des Zustandsspeichers

Da der Zustand Z_k vom Schaltnetz während der ganzen Taktphase gebraucht wird, muss man als Speicher flankengetriggerte D-Flipflops oder äquivalente Schaltkreise verwenden.

Realisierung des Schaltnetzes

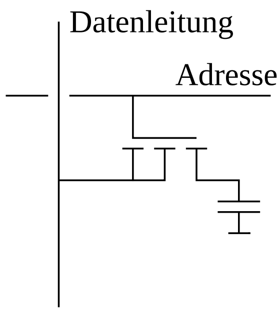
Das Schaltnetz kann im Prinzip mit den im letzten Kapitel besprochenen Methoden aus Gattern aufgebaut werden. Höhere Flexibilität erreicht man, indem man einen addressierbaren Speicher verwendet (Z_k und X bilden die Adresseingänge, Y und Z_{k+1} die Datenausgänge: *Lookup table*). Allerdings ist die Zahl der Eingangs- und Zustandsvariablen auf diese Weise sehr beschränkt (einige 20), während der grösste Teil des Speicherbereichs oft gar nicht verwendet wird. Deshalb entstand eine ganze Familie von speziellen programmierbaren Logikbausteinen, über die wir im übernächsten Abschnitt eine Übersicht gewinnen wollen.

Man unterscheidet zwischen *Tabellenspeicher* (*Memory*) und *Funktionsspeichern* (*programmable logic device*).

7.2 Tabellenspeicher

Tabellenspeicher sind dadurch gekennzeichnet, dass sie für *jeden* Eingangszustand (Adresse) Daten abspeichern und wieder ausgeben können. Sind n Adressbits vorhanden, können also 2^n Zustände gespeichert werden. Solche Speicher werden mit RAM (random access memory) bezeichnet. Oft werden mehrere Datenbits an einer Adressstelle gespeichert, sind es 8 spricht man von einem *Byte* organisierten Speicher, 4 bit heissen ein *nibble*. Sind es soviele Bits wie die zughörige Schaltung (zum Beispiel ein Computer) maximal brauchen kann, spricht man von *word* organisierten Speichern.

Man unterscheidet statische und dynamische RAM. Statische RAMs bestehen im Prinzip aus 2^n Flipflops in einer möglichst einfachen Form.



Dynamische Speicher (DRAM) benützen als Speicherzelle eine Kapazität. Das ist platzsparend, hat aber den Nachteil, dass die Speicherzelle nach jedem Lesezugriff wieder neu geladen werden muss (Ladungsverlust auf den Anschlussleitungen). Ausserdem muss die Ladung von Zeit zu Zeit durch Lesen und wieder Beschreiben erneuert werden (refresh cycle). Moderne DRAMs haben diese Logik bereits eingebaut. 27.1.99

FIFO und Ringspeicher

Unter einem *FIFO* (*first in – first out*) Speicher versteht man einen System, indas man Daten eingeben, und sie in der gleichen Reihenfolge wieder auslesen kann. Das ist an sich nichts besonders, jedes Schieberegister hat diese Eigenschaft.

Von einem FIFO spricht man allerdings erst dann, wenn die Auslese völlig asynchron zum Abspeichern passieren kann. FIFOs werden als Zwischenspeicher zwischen zwei an sich ansynchronen Systemen verwendet, zum Beispiel zwischen einem ADC (analog digital converter), der mit fester Abtastfrequenz ein Signal ausmessen soll, und dem Computer, der die entstehenden Daten von Zeit zu Zeit einlesen in grösseren Blöcken einlesen soll. Man spricht auch von einem *elasticity buffer*.

FIFOs werden als Ringspeicher realisiert. Ein ringförmiger Speicher der Tiefe n ist eine normaler Speicher, dessen Adressen durch einen Zähler bei jedem Schreib- bzw. Lesevorgang um eins erhöht werden. Wird n erreicht, geht es bei 1 weiter. Es gibt einen seperaten Schreib- und Lesezähler, die Read- und Writepointers A_r und A_w genannt werden.

Ein FIFO hat neben Datenein- und Ausgängen je einen Read und Writestrobe als Eingänge, mit denen man den Zeitpunkt für das Lesen und Schreiben bestimmen kann. Ausserdem gibt der FIFO je ein Signal 'FIFO full' ($A_w - A_r = n$), 'FIFO empty' ($A_r = A_w$) und 'FIFO half full' ($A_w - A_r \geq n/2$) von sich, die durch Vergleich der beiden Adresszähler entstehen. Die Speicherorganisation der FIFOs wird in der Form nxb angegeben, wobei n die Tiefe, und b die Breite angibt. Zum Beispiel hat der 74ACT3641 eine 1k x 36 Organisation. Das bedeutet er kann 36 bits auf einmal ein- oder auslesen und es ist $n = 1024$. Dieser chip hat 120 pins und läuft mit 50 MHz.

ROM

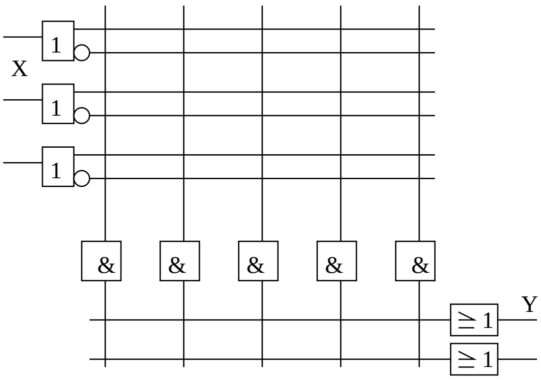
ROM sind Read Only Memories. Das heisst sie werden im normalen Betrieb nur gelesen, und nicht beschrieben. Natürlich müssen sie bei der Initialisierung 'einmalig' beschrieben werden können. Man unterscheidet:

- MROM, oft einfach ROM, Masken – ROM. Bei der Herstellung wird die letzte Metallisierungsschicht nach den Kundenwünschen speziell hergestellt. Sie enthält dann die programmierte Information.
- PROM, progammierbare ROM. Hier ist eine einmalige Programmierung durch einen speziellen Programmierpuls mit erhöhter Spannung möglich (zum Beispiel werden Kontakte oder Dioden durchgebrannt).
- EPROM, UV–EPROM, mit Ultraviolett – Strahlung löschbare ROMs. Hier besteht die Speicherzelle aus einem MOSFET, der ein zusätzliches 'floating gate' ohne externen Anschluss zwischen dem normalem gate Anschluss und dem Kanal hat. Die Programmierung erfolgt durch Aufladen dieses gates mit 'heissen' Elektronen, die durch Überspannung erzeugt werden. Mit Hilfe von Ultraviolettstrahlung können die Elektronen wieder befreit, der chip also gelöscht werden. EPROMs haben deshalb ein kleines Quarzfenster.
- EEPROM. Funktionsmässig wie EPROM, das Löschen erfolgt aber elektrisch über einen speziellen Lösch- und Schreibanschluss. Typische Lösch- und Schreibzeiten sind einige 10 ms.

Bei einem 'nonvolatile RAM' handelt es sich in der Regel um eine Kombination von einem gewöhnlichen RAM mit einem EEPROM. Bei Ausfall der Stromversorgung wird mit dem letzten Rest der Spannung, der Inhalt des RAM auf das EEPROM kopiert. Beim Einschalten wird als erstes der Inhalt des EEPROM in das RAM kopiert. Manchmal handelt es sich aber bei einem nonvolatile RAM einfach um einen Speicher mit Batterie...

7.3 Programmable Logic Devices

Man unterscheidet grundsätzlich drei verschiedene *PLD*. Nämlich *PAL* (*programmable array Logic*), *PLE* (*programmable logic element*) und *LCA* (*logic cell array*). PAL und PLE sind grundsätzlich gleich aufgebaut und stellen eine direkte Realisierung der disjunktiven Normalform einer Wahrheitstabelle dar. Das folgende Beispiel zeigt eine solche mit drei Eingängen X und zwei Ausgängen Y:



Dabei können beim PAL jeweils die Verknüpfungen der Eingänge zu den AND Elementen programmiert werden, während die OR – Zuordnungen am Ausgang festgelegt sind. Bei einem PLE sind die AND Zuordnungen fest (Adressen), während die OR – Zuordnung zu den Ausgängen programmierbar ist (Daten). Ein PLE ist also dasselbe wie ein PROM (!).

PALs werden mit Vorteil dann eingesetzt, wenn es viele Eingänge gibt, jedoch nur wenige verschiedene Kombinationen überhaupt betrachtet werden müssen. Ausserdem sind sie oft schneller in der Durchlaufzeit als PROM.

FPGA

LCA werden auch *FPGA* (*field progammable gate array*) genannt, und unterschieden sich grundsätzlich von den PALs dadurch, dass sie aus vielen einzelnen *CLB* (*configurable logic blocks*) bestehen.

Solche CLB bestehen zum Beispiel bei der XILINX 4000er Serie aus zwei D – Flipflops mit Set und Reset, deren D – Eingänge aus einer programmierbaren logischen Funktion mit 8 Eingängen gebildet werden können. Clock, Set und Reset können ebenfalls auf verschiedene Weise bedient werden. Für Details siehe [XILINX94], Seite 2-10. Es handelt sich also um eine Art Mini FSM.

Typische FPGAs bestehen aus 64 bis 1024 solchen CLBs. Die Verbindung zwischen den einzelnen CLBs ist durch in $x - y$ Form angeordnete Leitungen programmierbar. Ausserdem gibt es IOB (input output blocks), bei denen man programmieren kann, ob ein bestimmter pin input oder output sein soll, ob die Daten mit einem D-FF zwischengespeichert werden sollen und ob es sich extern um TTL oder CMOS Level handeln soll. (siehe Seite 2-14)

Ausserdem existiert ein globaler asynchroner Reset.

Das eingebaute Konfigurationsmemory ist als RAM realisert, und hat eine typische Grösse von 400 kBit. Man ladet es entweder von einem Computer, oder von einem PROM beim power up reset (Logik dafür ist im FPGA schon eingebaut).

Typische Anwendungen sind alle logischen Schaltungen, die flexibel und doch sehr schnell sein müssen (40 MHz sind problemlos, über 100 MHz möglich), und wo Geld keine Rolle spielt:

- Jede Art von endlichen Automaten
- Pipeline Architekturen: Wenn ein bestimmter Schritt eines Algorithmus fertig ist, werden die Resultate der Logik für den nächsten Schritt weitergereicht, und gleichzeitig neue Daten eingelesen.
- Ein 16 bit adder braucht 9 CLBs
- Ein 16 bit up/down counter braucht 8 CLBs

Für eine bestimmte Anwendung müssen die CLB und die Verbindungen optimal programmiert (geroutet) werden. Dafür stellen die Hersteller Software zur Verfügung. Dabei geht es nicht nur um die richtigen Verbindungen, es muss auch eine Simulation gemacht werden, um zu kontrollieren, ob das Timing stimmt (Beipiel FSM: Die clocks für die verschiedenen D-FF müssen hinreichend synchron kommen, Schaltnetz propagation delay muss $< 1/f_max$ sein).

Klarerweise sind wie immer in der digitalen Logik in der Regel die Zahl der Anschlusspins die begrenzende Grösse.

Als Eingabeform für das routing wird heute meist eine HDL verwendet, wie wir das im nächsten Kapitel an zwei Beispielen ansehen wollen. 1.2.99

Boundary Scan

Im Zusammenhang mit komplexen logischen Schaltungen stellt sich immer auch die wichtige Frage des Testens. Testen kann sehr schwierig sein, rein aus mechanischen Gründen (keine Anschlüsse und kein Platz für Testproben), aber auch aus GrÜnden der Komplexität. Es ist oft sehr schwierig, das richtige Testprozedere zu definieren.

Der sogenannte *boundary scan* soll das Testen erleichtern. Dabei werden alle IOB eines FPGA oder anderen logischen chips in einem grossen Schieberegister miteinander verbunden. Man kann dann ein Testpattern seriell in die IOB schieben, den chip fuer eine wohldefineirte Anzahl Taktzyklen einschalten und anschliessend das Resultat wieder über das gleiche Schieberegister zurücklesen. Anschlüsse und Bedienungsalgorithmus von boundary scans sind international genormt.

7.4 Hardware Description Languages

HDL (*hardware description languages*) sind logische Sprachen, die dazu dienen, eine digitale Schaltung algorithmisch zu beschreiben. HDL Texte sehen ähnlich aus wie gewöhnliche Computersprachen (z.B. c, FORTRAN). Sie sind mit einigen Elementen ergänzt, die es erlauben, den zeitlichen Ablauf der Systeme richtig zu beschreiben.

Logische Schaltungen werden heute fast immer in HDL formuliert. Dies erlaubt:

- eine eindeutige Beschreibung mit festen Regeln, was die Schaltung eigentlich können soll. Der HDL Compiler zeigt, ob der Algorithmus in sich konsistent ist.
- eine Simulation des Algorithmus mit beliebigen Eingangsbedingungen.
- eine direkte Synthese der Schaltung als 'IC–Sarg', für FPGA, für andere PLDs, oder eine direkte ASIC Produktion mit Hilfe von Standardlibraries.
- Eine weitere Simulation erlaubt das detaillierte Studium des Verhaltens der Implementation. Hier kann vor allem auch das timing kontrolliert werden.
- Die autmatische Erzeugung von Testvektoren samt zu erwartendem Ausgang für spätere Hardwaretests.

Synthese ist in der Regel allerdings nur von einer eingeschränkten Mächtigkeit der HDL aus möglich (sogenannter RTL, Register Transfer Level, wo die Architektur festliegt: Was sind die Zustandsvektoren, FSM, Zähler).

Beispiel Kaffeemaschine (in VHDL)

Beispiel: 16 bit counter in Verilog

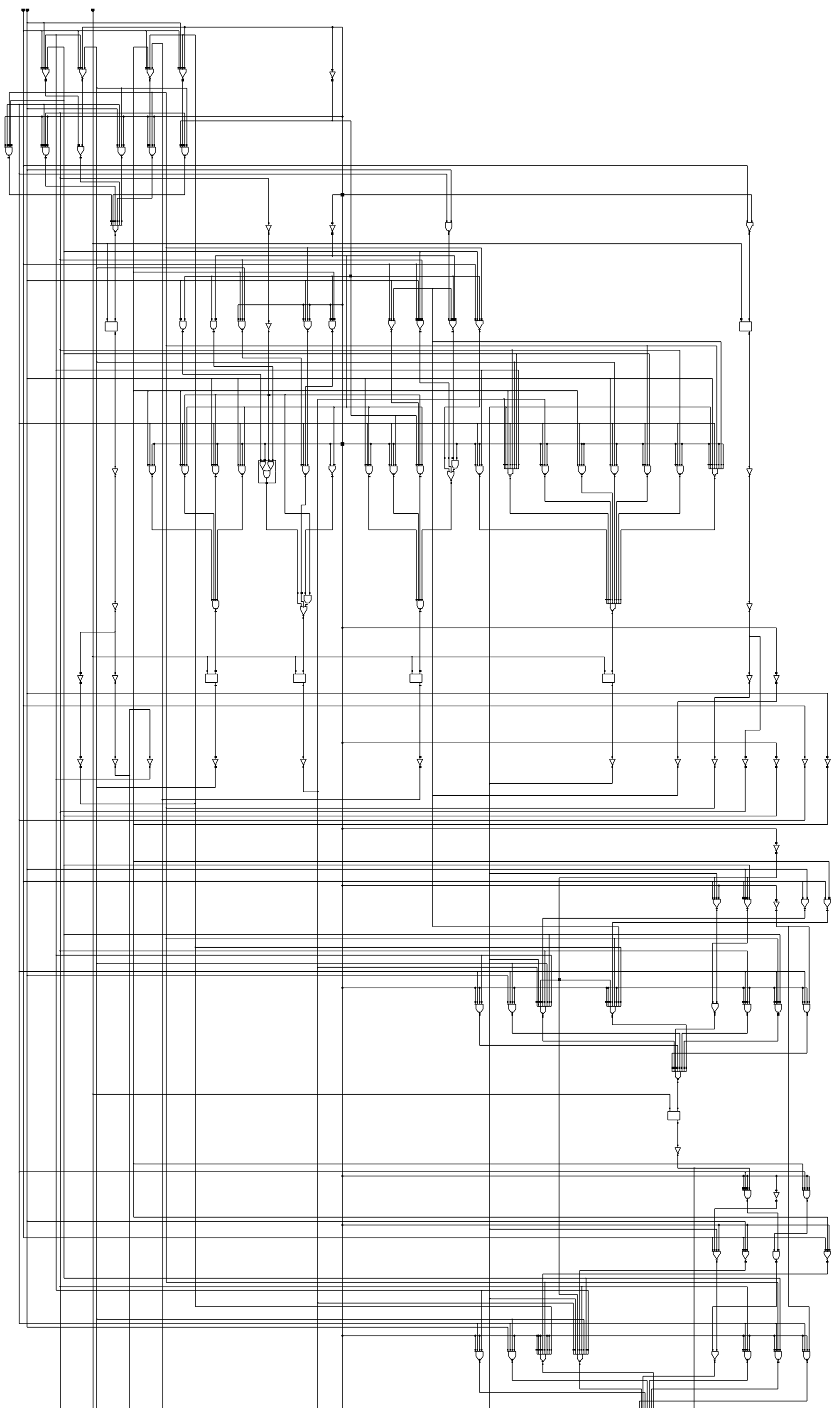
Das folgende Beispiel hat uns U. Trunk vom MPI Heidelberg zur Verfügung gestellt. Vielen Dank!

Die Verilog – HDL Beschreibung des 16 bit up/down counters mit synchronem Reset lautet:

```
module Counter (
        pulse ,   //cout pulse input
        dir,      //direction flag 1=up 0=down
        reset,    //reset
        data);    //counter data outputs (8bit)

input   pulse,
        dir,
        reset;
output  [7:0] data;

reg     [7:0] data;



always @(posedge pulse)
   if (reset == 1)
     data[7:0] = 8'b00000000;  //synchronous reset
   else
     if (dir == 1)             //counting up or down....
       data = data + 1;
     else
       data = data - 1;

endmodule
```

Der Compiler stellt fest, dass alles konsistent ist. EIne Simulation lieferte die erwarteten Resultate. Schliesslich produziert ein Synthese Programm das folgende Schaltbild auf der nächsten Seite.

Die Implementation dieser Schaltung in einen ASIC mit Hilfe einer Standarbibliothek von Transistoren und digitalen Grundschaltungen wird schliesslich schon etwas unübersichtlich. (siehe Overheadfolie)



Kapitel 8

Signalübertragung

Die Übertragung elektrischer Signale über endliche Strecken ist einer Vielzahl von Schwierigkeiten ausgesetzt, die in der Praxis zu den häufigsten Problemen mit elektronischen Systemen führen.

- Lange Leitungen ergeben Laufzeiten, Reflexionen, Dämpfung und Dispersion. Eine Leitung muss dann als lang gelten, wenn 1/Anstiegszeit der Signale kleiner als die halbe Leitungslänge ist (siehe Kapitel 1.5).
- Die Kapazität der Leitungen und der Anschlüsse (Stecker, Eingangskapazitäten der Empfängerbausteine etc.) erzeugen einen Umladestrom, der umso grösser wird, je steiler die Anstiegszeiten sind. Dieser Strom fliesst auch in den Masseverbindungen.
- Ströme in den Massenverbindungen ergeben durch ihren ohm'schen Widerstand eine zeitlich variable Potentialdifferenz zwischen Sender und Empfängerbaugruppen (*ground shift potentials*, *ground bouncing*).
- Signale in parallel laufende Leitungen erzeugen Übersprechen (*Crosstalk*, *X-talk*), das heisst die Signale sind nicht nur in ihrer eigenen Leitung sichtbar, sondern auch in benachbarten.
- Die elektromagnetische Umweltverschmutzung führt dazu, dass Leitungen auch Signale aufnehmen, die von anderen Systemen elektromagnetisch abgestrahlt wurden (*Pickup*).

Diese Probleme sind natürlich in erster Linie für die schwachen analogen Signale von Sensoren relevant. Es zeigt sich aber, dass auch bei schnellen digitalen Systemen diese Überlegungen relevant sind. Dabei treten sie nicht nur bei Kabeln, sondern auch bei Verbindungen auf einem p.c.b, oder manchmal sogar innerhalb eines chips auf. Besonders kritisch sind dabei 'mixed signal' Elemente, also chips, die gleichzeitig analoge und digitale Signale verarbeiten.

8.1 Digitale Signalstandards

Die bisher diskutierten logischen Standards TTL und CMOS eignen sich beide nicht für schnelle Signalübertragung. Die Bausteine haben in der Regel nicht genügend Leistung um den Strom für ein 50 oder 100 Ω Kabel zu liefern. Die Rauschempfindlichkeit der Eingänge ist viel zu hoch (z.B. 0.4V bei TTL), sodass X-talk, Pickup und Ground-bouncing zu Fehlübertragungen führen. Um den X-talk klein zu halten, möchte man langsame Anstiegszeiten und kleine Amplituden, was gerade eine gegenteilige Zielsetzung gegenüber den Standard - TTL und CMOS Bausteinen darstellt.

Man unterscheidet *single ended transmission* (oder *unbalanced transmission*) und *differential transmission* (oder *balanced transmission*), je nachdem ob nur eine Leitung vorhanden ist, und der Signalstrom durch die Masseverbindung zurückfliesst, oder aber of eine richtige differentielle Doppelleitung vorhanden ist. Maximale Übertragungsrate und Leitungslänge sind jeweils die wichtigsten Kenngrössen.

Im weiteren werden unter anderem Anzahl Sender und Empfänger, Spannungshub beim Sender und minimaler Spannungsdiffereny am Empfänger, minimale Leitungsimpedanz und Kurzschlussfestigkeit definiert.

Im folgenden werden die häufigsten Standards vom elektronischen Standpunkt aus behandelt. Die logischen Aspekte werden im Computerpraktikum diskutiert werden.

8.1.1 IEEE 488

Beim *IEEE488* oder *GPIB* handelt es sich um einen 8 bit breiten Bus, mit 8 zusätzlichen Kontrollleitungen mit dem TTL Standard. Er wird für die Verbindung von Mess- und Kontrollgeräten mit Computern verwendet. Es muss genau ein Buscontroller pro Bus vorhanden sein, es sind aber mehrere aktive Sender möglich (open collector Techink). Sein Vorteil besteht in seiner weiten Verbreitung, die unter anderem in der sehr vollständigen Normierung, inklusive genaue Bedeutung der Kontrollsignale, erlaubte Lasten, Steckertyp und -Belegung usw. begründet ist.

Die maximale Datenübertragung beträgt 1.5 MByte/sec. über maximal einige Meter. 3.2.99

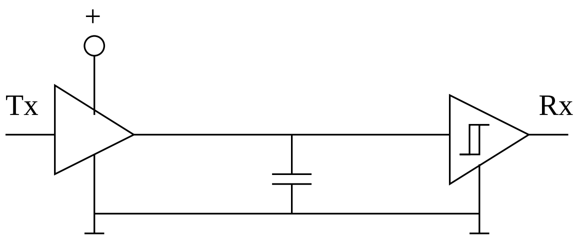
8.1.2 EIA-232

Dabei handelt es sich um eine sogenannte 'single ended transmission'. Das heisst das Signal wird über *eine* Leitung übertragen, der Rückstrom muss über die Masseleitung erfolgen.

Dieses System hat den Nachteil, dass Massepotentialdifferenzen entsprechende Übertragungsfehler erzeugen können. Die Übertragungsrate ist auf etwa 20kBit/sec in 20m Distanz beschränkt.

Dafür ergeben sich geringe Kosten und einfache Implementierungen, sodass dieses System immer noch sehr verbreitet ist.

Die Signale müssen mindestens ± 3V Amplitude haben (positiv heisst '0', negativ '1') und sollen eine Anstiegszeit haben, die *kleiner* als 30V/μs bleiben soll, damit es kein X-talk gibt.



Die maximale erlaubte Kabelkapazität ist mit 2500 pF spezifiziert, was zu einer Leitungslänge von typisch 20 m führt. Diese Daten zusammen mit der max. Übertragungsrate von 20 kbit/sec führen dazu, dass keine Abschlusswiderstände nötig sind.

DIe EIA-232 Norm beschreibt auch einen 25-poligen Stecker mit diversen Kontrollleitungen. Oft wird auch ein 9-poliger DB9S Stecker verwendet, der ein subset der EIA-232 Signale enthält. Die DTE - Seite (data terminal equipment, Stiftstecker, z.B. PC) wird mit der DCE (data communication equipment, Buchsenstecker, z.B. Modem) verbunden. Die Bedeutung der Leitungen sind

- Pin 1, DCD, data carrier detect. von DCE nach DTE. Aktiv, wenn das Modem gültige Daten empfängt.
- Pin 6, DSR, data set ready. von DCE nach DTE. Aktiv, wenn das Modem Verbindung hat.
- Pin 2, RD Receive Data. von DCE nach DTE.
- Pin 7, RTS, request to send, von DTE nach DCE
- Pin 8, CTS, clear to send, von DCE nach DTE
- Pin 3, TD, Transmit data, von DTE nach DCE

- Pin 4, DTR, data terminal ready, von DTR nch DCE
- Pin 9, RI, Ring indicator, von DCE nach DTR
- Pin 5, Masse

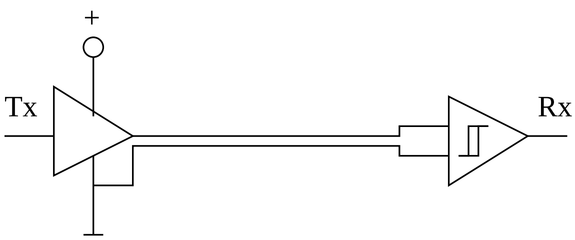
Will man zwei DTE, z.B. PC's miteinander verbinden, so braucht es ein Nullmodem, dass vor allem die Pins 2 und 3 vertauscht, sowie die verschiedenen Kontrolleitungen richtig verbindet (je 7,8 mit 1 und 4 mit 6 verbinden).

Die Datenübertragung wird mit einer Flanke von + nach - auf der Datenleitung eingeleitet. Nach einem Takt (Startbit, '1') folgen 7 oder 8 Datenbits, evt. gefolgt von einem Parity bits und einem oder 2 Stopbits ('0'). Eine solche Datengruppe wird *frame* genannt. Man beachte, dass keine separate clock übertragen wird (*asynchronous transmission protocoll* ATP). Die EIA-232 Leitungen werden mit UART chips (*universal asynchronous Receiver and Transmitter*) bedient.

Die heute meist gebräuchliche Einstellung ist 8N1, das heisst 8 bits, keine Parity und 1 Stopbit. Natürlich muss auch die Rate auf beiden Seiten richtig eingestellt sein. Zum Teil sind die Systeme in der Lage die Taktfrequenz zu erkennen.

8.1.3 Differentielle Systeme: RS423, RS 485 und LVDS

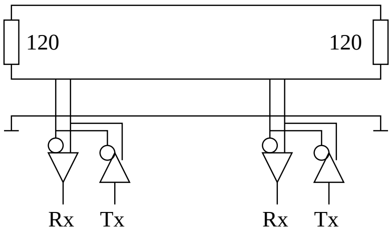
Verwendet man auf der Empfängerseite einen Differenzverstärker, und führt jeweils je eine Daten- und Masseleitung separat, kann man das Problem der Potentialdifferenzen weitgehend umgehen. Ebenfalls wird der pickup wegsubtrahiert.



Die RS423 Norm erlaubt Datenübertragungen von 120 kbps (3 kbps) über Distanzen von 30 m (1200 m). Es dürfen maximal 10 Empfänger an einen Sender angeschlossen werden.

Die Mindestpulshöhe beträgt $\pm$200mV. Die Eingangsimpedanz beträgt 4kΩ , es gibt also noch keinen Wellenabschluss des Kabels.

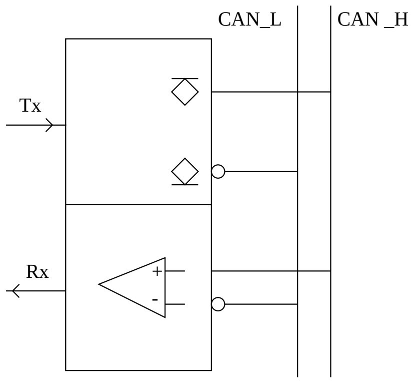
Geht man schliesslich diesen Weg konsequent weiter, kommt man zum RS 485 (ISO 8482) Standard, der aus einem vollständig differentiellen 120 Ω Leitung mit korrektem Abschluss an beiden Enden besteht. Es sind nun mehrere Sender möglich. Die Eingangsimpedanz der Empfänger beträgt 12kΩ , es können maximal 32 Empfänger oder Sender angeschlossen werden.



Die Schwellwertspannung am Empfänger beträgt ±200 mV, die typische Signalemplitude ist 3V, was zusammen mit den 100 Ω Leiterimpedanz eine erhebliche Verlustleistung ergibt. RS485 Treiber sind oft thermisch überlastgeschützt.

Man kann bis zu 10 MBit/sec und bis über 1000 m Daten übertragen. Allerdings sind die Interfacebausteine teurer, und man muss twisted pair Kabel nehmen. Beachte, dass die Masseverbindung nötig ist, um sicherzustellen, dass die Signale am Empfänger im erlaubten Spannungsbereich (-7V ... +12V) bleiben.

Das *CAN* (*controller area network*) funktioniert im Prinzip mit einem solchen Bus, wobei eine bus arbitration so definiert ist, dass ohne Kontrollleitungen viele Geraete am gleichen Bus hängen koennen.



Die Treiber der Leitungen sind als open collector ausgeführt, sodass ein logische '1' dem passiven Zustand entspricht. Damit kann ein Sender testen, ob der Bus gerade frei ist. Dafür wird vorerst ein 12-bit langes bus arbitration Wort gesendet. Sendet er eine '1', und sieht aber eine '0', so weiss der Sender, dass ein anderer Sender aktiv ist und stellt sich selber wieder ab. Dieses Verfahren ist im Detail festgelegt, und erlaubt ein volle Busfunktionalität mit einem minimalen Verkabelungsaufwand.

High Speed CAN kann eine Datenrate von 1 MBit/s über eine Distanz von 40 m übertragen. Maximal 30 Geräte können an der Leitung mit 120 Ω IMpedanz hängen, Abzweigungen dürfen maximal 30 cm lang sein. Typischer Driver chip: SN75LBC031. Nachdem CAN ursprünglich in der Automobilindustrie erfunden wurde, hat es in jüngster Zeit eine starke Verbreitung im Bereich der Kontrollsysteme gefunden.

Das *LVDS* (low voltage differential system) ist eine relativ neue Entwicklung und beruht auf demselben Prinzip, arbeitet aber mit kleineren Spannungen (Schwellwert 100 mV, Amplitude typisch 350 mV), sodass Übertragunsraten bis mindestens 155 Mbps möglich sind. Hier wird bei den Treibern der Strom reguliert (typisch 3.5 mA), sodass die Abschlusswiderstände auf jeden Fall gebraucht werden. Ein typischer Vertreter ist der DS90C031 (Sender), DS90C032 (Empfänger). Letzterer hat ausserdem die angenehme Eigenschaft, dass offene Eingänge einen stabilen HIGH Zustand am Ausgang garantieren.

LVDS ist zu empfehlen für alle neue gebauten Systeme. Es ergibt die kleinsten X-talk Probleme und die grösste Störsicherheit. Man erreicht die bis zu 200 Mbps über 2 m, oder 5 Mbps über 100 m.

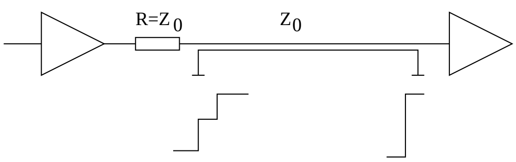
8.2 Pickup, X-talk und Rauschen in der Praxis

Bisher haben wir bei verschiedenen Gelegenheiten diverse unerwünschte Effekte bei der digitalen (und analogen) Datenübertragung kennengelernt: Über- und Unterschwinger wegen Reflektionen, X-talk, pickup, ground bouncing. Die folgenden Hinweise und Regeln sollen mithelfen, diese Probleme in der Praxis zu minimieren:

Überschwinger

Ein korrekter Leitungsabschluss hilft *Reflexionen* zu vermeiden. Leider ist das aber nicht immer möglich (DC Strom reicht nicht, Stecker und Kabelinhomogenitäten produzieren Reflexionen, Abzweigungen an Kabeln können nicht immer kurz genug gehalten werden, usw.)

Für nicht zu hohe Übertragungsfrequenzen ($f \ll 1/\text{Laufzeit}$) ergibt das Anpassen des Ausgangsimpedanz des Treibers an die nominelle Kabelimpedanz oft eine befriedigende Lösung:



Das Signal auf der Senderseite geht vorerst entsprechend dem Spannungsteiler aus R und dem Kabel auf die halbe Spannung. Wenn die reflektierte Welle zurückkommt, wird die volle Spannung erreicht.

Auf der Empfängerseite gibt es in erster Näherung weder Über- noch Unterschwinger. Diese Schaltung mit offener Leitung eignet sich auch gut, um Impedanzen von Leitungen zu messen: Aus dem Verhältnis der ersten zur zweiten Amplitude bestimmt man Z_0, aus der Laufzeit daraus L und C der Leitung.

Ausserdem werden oft auf der Empfängerseite Dioden in Sperrrichtung gegen Masse und Stromversorgung eingebaut. Dies vernichtet die Energie der Überschwinger.

Übersprechen

Laufen zwei Leitungen parallel so ergibt sich eine gegenseitige Induktivität L_c und eine solche Kapazität C_c. Diese werden analog der Leitungsimpedanz in eine komplexe gegenseitige Impedanz $Z_c = \sqrt{L_c/C_c}$ (*coupling impedance*) zusammengefasst: Der *Crosstalk*

C ergibt sich dann:

$$C := \frac{V_2}{V_1} = \frac{1}{1 + 2\frac{Z_c}{Z_0}} \tag{8.1}$$

wobei V_2 die Amplitude des induzierten, V_1 diejenige des ursprünglichen Signales und Z_0 die Leitungsimpedanz bedeutet.

Für printed circuits zeigt die folgende Tabelle drei Beispiele für gegenseitige Impedanzen und Crosstalk. Es sind parallel laufende Leiterbahnen angenommen, deren Breite gleich dem Abstand zwischen den Bahnen ist. Verglichen werden die Fälle, ohne und mit Massenebene auf der Rückseite ("Backplane"), sowie der Fall, bei dem eine Leiterbahn zur Abschirmung ihrer Nachbarn voneinander auf Masse gelegt ist:

	Z_0 [Ω]	Z_c [Ω]	C [%]
ohne backplane	200	100	50
mit backplane	80	125	25
mit Masse zwischen den Leiterbahnen	100	400	11

10% crostalk ist in digitalen designs durchaus üblich und in der Regel funktionieren die Schaltungen problemlos.

Analysiert man die Sache genauer (siehe zum Beispiel im [TI-98], page 2-23 bis 2-27), findet man, dass der sogenannte "backward crosstalk", also das übersprechende Signal am Anfang der Leitung stärker ausgeprägt und positiv ist und durchaus das mehrfache an Amplitude erreichen kann. Es ist deshalb wichtig, dass parallele Leitungen mit crosstalk an beiden Enden richtig abgeschlossen werden! Der "forward crosstalk" ist in der Regel negativ.

Für Kabel wird der X-talk im allgemeinen viel grösser. Deshalb werden twisted pair Kabel verwendet, bei denen man ebenfalls etwa 10% x-talk erreicht.

Pickup

Unter *Pickup* versteht man das Einkoppeln von fremden Signalen auf die Datenleitungen. Eigentlich ist der X-talk eine spezielle Form des Pickups.

Beim elektrischen Einkopplen wirkt die kapazitive Kopplung C_c zwischen der Störquelle und der Übertragungsleitung. Das eingekoppelte Signal ist demnach differenziert und proportional zu RC_c, wo R der Abschlusswiderstand der Leitung ist (falls $f \ll 1/RC_c$). Gegen elektrisches Einkoppeln hilft also eine elektrostatische Abschirmung ("Faradaykäfig") und eine kleine Eingangsimpedanz der Empfänger.

Magnetische Einkopplung entsteht, wenn in einer Leiterschlaufe durch das Fara-dy'sche Induktionsgesetz eine Spannung induziert wird. Dagegen hilft das Vermeiden von Leiterschlaufen, insbesondere muss darauf geachtet werden, dass Masseleitungen schön ordentlich parallel zur Signalleitung geführt wird (am besten twisted pair). Weitere Masseleitungen sind nach Möglichkeit zu vermeiden.

Schliesslich spricht man von galvanischer Einkopplung, wenn der Störstrom direkt auf die Signalleitung oder die zugehörige Referenzmasse Verbindung hat.

Abschirm- und Erdungsregeln

Die folgenden *Abschirmregeln* und die *Erdungsregeln* gelten natürlich insbeondere auch für analoge Signale. Besonders zu beachten sind sie in der Gegend von physikalischen Sensoren.

1. Elektrische Abschirmungen von Kabeln sollen immer auf beiden Seiten angeschlossen werden, da sonst das freie Ende als Antenne wirkt, und entsprechende Signale in die Signalleitungen einkoppelt (einseitig hilft manchmal bei kleinen Frequenzen in der Audiotechnik).

2. Zusätzliche Masseleitungen sollen eng parallel, wenn möglich verdrillt zu den Signalkabeln geführt werden (keine Leiterschlaufen).

3. Jede Art von Strömen auf den Masseleitungen vermeiden. Alle Signale differentiell übertragen. Niemals Stromversorgungsströme über Referenzmasse fliessen lassen!

4. Mögliche Rauschquellen müssen ebenfalls gut abgeschirmt werden.

5. Abschirmungen müssen "wasserdicht" sein, Lochgitter nützen überhaupt nichts (Durchgriff des elektrischen Feldes), sie müssen dick genug sein, sodass einerseits der ohm'sche Widerstand klein gegenüber allen Signalleitungsimpendanzen ist, und die Dicke gross gegenüber der Skintiefe ist (Beispiel: Alu bei 100 MHz: $\delta = 8.5\mu$m, bedampfte Mylarfolie hilft also nichts, aber Haushaltsfolie ist oft o.k.).

6. Alle leitenden Teile innerhalb der Abschirmung müssen an *einem* festen Punkt sternförmig angeschlossen sein. Hier sollen ebenfalls die Abschirmungen der Zuführungsleitungen angeschlossen sein.

7. Bei Sensoren stets den gesamten Signalkreislauf betrachten. Darauf achten, dass dieser so kurz als möglich ist, keine fremden Ströme darin fliessen und keine Leiterschlaufen vorhanden sind.

In kritischen Fällen hilft oft eine doppelte Abschirmung, wobei eine innere Abschirmung nur gerade den Sensor und die erste Vorverstäkerstufe umfasst. Diese ABschirmung wird mit einem möglichst kurzen Kabel an die Masse dieses Verstärkers angeschlossen. Eine äussere Abschirmung umfasst alle Teile des Systems. Die Signal- und Stromversorgungsleitungen sollen nur an einer einzigen Stelle gemeinsam durch ein Loch in das innere gelangen. Am Rand dieses Loches wird der "Massenstern" angebracht, an dem die Abschirmung der äusseren Kabel, die Stromversorgungsmasse, die äussere Abschirmung und alle übrigen leitenden Teile des Systems direkt (sternförmig) angeschlossen werden.

Kapitel 9

Datenakquistionssysteme

Datenakquisitionssysteme dienen dazu, physikalische Messdaten zu erheben, und sie dauerhaft auf elektronischen Rechenanlagen abzuspeichern. Sie bestehen aus folgenden Elementen:

- Der *Mesfühler* oder *Sensor* dient der Umwandlung der zur messenden physikalischen Grösse in einen elektrischen Strom oder Spannung. (siehe Kapitel 5).
- Ein Verstärker unmittelbar neben dem Sensor wird benötigt, um die in der Regel kleinen Ströme oder Spannungen in grössere Spannungen umzuwandeln, sodass sie störungsfrei über eine grössere Distanz übertragen werden können.
- Ein analoges Filter erlaubt eine erste Verarbeitung des Signales. Im einfachsten Falle handelt es sich um ein Antialiasfilter (siehe übernächster Abschnitt), es können aber auch kompliziertere Funktionen interessant sein (z.B. peakfinding etc.). Hier müssen alle Funktionen eingebaut werden, die vor der Digitalisierung durchgeführt werden müssen.
- Ein ADC dient der Digitalisierung der Daten (siehe unten)
- Ein FIFO dient als Zwischenspeicher der Daten. (siehe 7.2).
- Ein digitales Rechensystem verarbeitet und speichert die Daten.

Analog – Digital – Wandler

ADC dienen der Umwandlung von analogen Spannungen in digital dargestellte Zahlen. Hier wollen wir nicht auf den internen Aufbau der ADC's eingehen, sondern nur deren charakteristischen Parameter besprechen. Generell gilt, dass ADC's mit umso kleineren maximalen Frequenz konvertieren können, desto genauer (höhere Zahl von Bits) sie sind.

Folgende Parameter sind für die Auswahl eines ADC's wichtig:

- Die maximale Taktfrequenz. Alle ADC's haben einen Takteingang, dessen positive oder negative Flanke den Zeitpunkt bestimmt, wann der analoge Spannungswert digitalisiert wird.
- Der maximale Eingangsspannungsbereich (full scale V_{fs}). Sie liegt im Bereich von 0.1 bis 10 V. Kleinere Spannungen müssen zuerst verstärkt werden, aber nicht zu nahe am ADC, wegen Clockpickup!
- Die Auflösung. Man gibt die Anzahl Bits n der digitalen Darstellung, oder der kleinste messbare Unterschied $\Delta V = V_{fs}/2^n$ an. Die Tatsache, dass der kleinste digitale Schritt immer endlich ist, führt zu einem Messfehler, den man auch den *Quantisierungsfehler* nennt. Entsprechend wird für die Abweichung vom wahren Signal der Begriff *Quantisierungsrauschen* verwendet.
- Die Darstellung von negativen Zahlen: Two's Complement (siehe 6.4.1) oder Binary Offset (das heisst der Nulpunkt der Eingangsspnanung wird auf den Mittelwert des digitalen Zahlenbereiches abgebildet).
- Die *aperture time* ist die Zeit, in der das Eingangssignal einigermassen konstant sein soll (Änderung $< \Delta V$). Sie ist in der Regel nur wenig ns lang.

Aliasing

Die endliche Abtastfrequenz f_c eines ADC's führt dazu, dass hohe Engangssignalfrequenzen f_s auf der digitalen Seite nicht mehr richtig erkannt werden können.

Man überzeugt sich leicht, dass für eine harmonisches Signal mit $f_s = f_c/2$ gerade noch die richitge Frequenz auf der digitalen Seite erkannt werden kann. Ist gerade $f_s = f_c$ erkennen wir überhaupt kein Signal mehr, der digitale Ausgang stellt eine Gleichspannung dar. Dazwischenliegende Frequenzen werden an $f_c/2$ gespiegelt.

Noch höhere Frequenzen werden - evt. gespiegelt - in den Frequenzbereich $[0 \ldots f_c/2]$ abgebildet. Dieser Effekt heisst *Aliasing*, und kann sehr verwirrend sein. Bei Verwendung digitaler Oszilloskope tritt er ebenfalls häufig auf. Er kann auch als Schwebung zwischen f_s und f_c interpretiert werden. Man spricht auch vom *Nyquisttheorem.*

Ein *Antialiasfilter* vor dem ADC Eingang wird als Tiefpass mit einer Grenzfrequenz unter $f_c/2$ und einer möglichst hohen Ordnung realisiert. Es sorgt dafür, dass kein Aliasing auftreten kann.

Index

Literaturverzeichnis

[Best87] Roland Best *Handbuch der analogen und digitalen Filterungstechnik*, AT Verlag Aarau, 1987

[Bron98] I.N. Bronstein, K.A. Semendjajew, G. Musiol, H. Mühlig: *Taschenbuch der Mathematik*, Harri Deutsch 1998.

[Ebel78] T. Ebel: *Regelungstechnik*, Teubner Studienskripten Nr. 57, 1978

[Hin96] Hermann Hinsch: *Elektronik, ein Werkzeug für Physiker*, Springer 1996.

[Horo97] Paul Horowitz, Winfried Hill *The Art of Electronics*, Cambridge University Press 1989, reprinted 1997.

[Jack90] Jackson, *Electrodynamics*

[Kori98] Ralf Kories, Heinz Schmidt-Walter: *Taschenbuch der Elektrotechnik*, Harri Deutsch 1998.

[Laker94] K.R. Laker, W.M.C. Sansen: *Design of analog integrated circuits and systems*, McGraw Hill 1994.

[Nühr98] Dieter Nührmann: *Das grosse Werkbuch Elektronik*, 4 Bände, Francis 1998.

[Pier96] Robert F. Pierret: *Semiconductor Device Fundamentals*, Addison-Wesley 1996.

[Putz71] W. Putz et al.:*Elektrotechnik und Kerntechnik*, Rowohlt Technik Lexikon, Hamburg 1971

[TI-98] E. Haseloff, H. Beckemeyer, J. Zipperer: *Data Transmission Design Seminar*, Texas Instruments 1998.

[Tiet91] U. Tietze, Ch. Schenk: *Halbleiter-Schaltungstechnik*, Springer 1991.

[XILINX94] XILINX, The programmable logic data book, San Jose, Ca. 1994.