

Herleitung der wichtigsten Eigenschaften von Diode, JFET und MOS Transistor

© Prof. Dr. Peter Fischer, ZITI, Uni Heidelberg

In diesem Text sollen einige wichtige Eigenschaften der p-n-Diode und des MOSFET hergeleitet werden. Zusätzlich wird der JFET beschrieben, da die Modulation des Drainstroms hier sehr anschaulich ist. Die Herleitungen kommen mit einem Minimum an Vorkenntnissen aus der Physik aus. Die wichtigsten Schritte und 'Zutaten' sind am Rand markiert. Dieser Text ersetzt nicht ein 'seriöses' Buch über Bauteilphysik, wie z.B. den Klassiker *S.M. Sze, Semiconductor Devices: Physics and Technology, Wiley, 1985, ISBN 0-471-87424-8*.

1 Die p-n-Diode

Zur Untersuchung der Eigenschaften der pn-Diode betrachten wir eine scharfe Grenzfläche zwischen n- und p-dotiertem Silizium. In der Praxis sind die Dotierungsprofile nicht so abrupt, dieser einfache Fall lässt sich aber leichter berechnen. Das p-Material mit einer Dotierungsdichte von N_A (Akzeptoratomen pro Volumeneinheit) befindet sich links, das n-Material mit einer Donatordichte N_D rechts. Dies ist in Abb. 1 skizziert.

1.1 Entstehung der Raumladungszonen

Wegen der ungleichen Konzentration der beweglichen Ladungsträger *diffundieren* Elektronen vom n-Material ins p-Material (\leftarrow) und Löcher vom p- ins n-Material (\rightarrow). Wegen der negativen Ladungen der Elektronen tragen *beide* Ladungsträger zu einem *Strom nach rechts* bei. Die abwandernden Elektronen und Löcher hinterlassen Bereiche ohne bewegliche Ladungsträger, die *Verarmungszone*, in der sich raumfeste, *geladene* Atomrümpfe befinden. Man spricht daher auch von einer *Raumladungszone* oder von der *Sperrschicht*. Der n-Bereich ist positiv geladen, der p-Bereich negativ. Dies führt zu einem elektrischen *Feld*, das in dieser Anordnung von rechts nach links zeigt, also negativ ist. Löcher *driften* in diesem Feld nach links, Elektronen nach rechts. Beide Ladungsträger tragen so zu einem *Feldstrom* bei, der von rechts nach links gerichtet ist. Im stationären Fall sind die Beträge dieser beiden Ströme gleich. Dies nutzen wir zur Berechnung der *Diffusionsspannung* aus, die sich durch den beschriebenen Ausgleichsvorgang ausbildet. Zur Berechnung legen wir den Nullpunkt der nach rechts gerichteten x-Achse auf die Grenzfläche. Die Raumladungszone im n-Bereich endet bei $x_n > 0$, die im p-Bereich bei $x_p < 0$ (s. Abb. 1).

Diffusionsstrom \rightarrow

Feld \leftarrow

Feldstrom \leftarrow

1.2 Berechnung der Diffusionsspannung

Wir beschreiben die Anordnung durch das ortsabhängige elektrische Feld $E(x)$ und durch die ortsabhängigen Dichten der Ladungsträger. Wir betrachten hier nur die Löcherdichte $p(x)$, die Rechnung für die Elektronendichte $n(x)$ verläuft analog.

Der Feldstrom ist proportional zum Feld, zur Dichte der Ladungsträger, zu deren Beweglichkeit μ (von der wir vereinfachend annehmen, daß sie nicht vom Feld abhängt) und der Ladung q :

Feldstrom

$$j_{\text{Feld}}(x) = q \mu p(x) E(x) \quad (1)$$

Der Diffusionsstrom hängt *nicht* vom Feld ab. Ladungsträger wandern von Stellen hoher Konzentration in Richtung niedriger Konzentration, also *entgegen* dem Gradienten der Ladungsträgerkonzentration, daher das Minuszeichen. Die Stärke der Diffusion wird durch die Diffusionskonstante D beschrieben:

Diffusionsstrom

$$j_{\text{Diff}}(x) = -q D \frac{dp(x)}{dx} \quad (2)$$

Im stationären Gleichgewicht (es fließt kein Strom durch die Diode) müssen sich der

Gleichgewicht

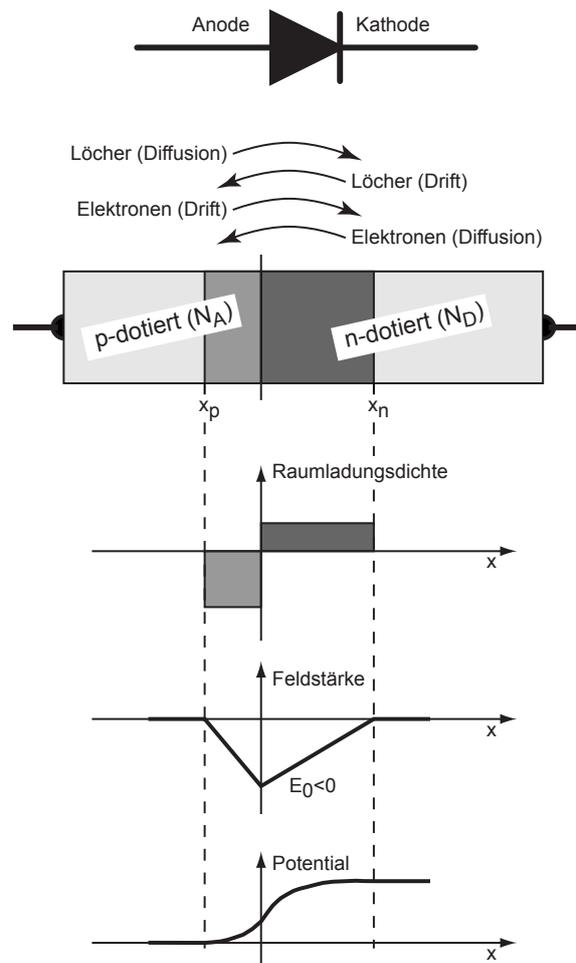


Abbildung 1: p-n-übergang, Raumladungszone, Feld und Potential

Feldstrom und der entgegengerichtete Diffusionsstrom aufheben, also

$$\begin{aligned} j_{\text{Feld}}(x) &= -j_{\text{Diff}}(x) \\ q \mu p(x) E(x) &= q D \frac{dp(x)}{dx} \\ \frac{\mu}{D} E(x) &= \frac{1}{p(x)} \frac{dp(x)}{dx} \end{aligned} \quad (3)$$

$$-\frac{q}{kT} \frac{dV(x)}{dx} = \frac{1}{p(x)} \frac{dp(x)}{dx} \quad (4)$$

$$-\frac{q}{kT} dV(x) = \frac{dp(x)}{p(x)}. \quad (5)$$

Im Schritt von (3) \rightarrow (4) wurde das Feld $E(x)$ durch den negativen Gradienten des Potentials ausgedrückt und die (nichttriviale!) Einstein-Beziehung $\frac{D}{\mu} = \frac{kT}{q}$ benutzt. In (5) wurde zu einem Differential übergegangen, das direkt integriert werden kann:

Einstein-Beziehung

$$\begin{aligned} -\frac{q}{kT} \int_{V_p}^{V_n} dV(x) &= \int_{p_p}^{p_n} \frac{dp(x)}{p(x)} \\ -\frac{q}{kT} (V_n - V_p) &= \ln \left(\frac{p_n}{p_p} \right) \\ V_{bi} := V_n - V_p &= \frac{kT}{q} \ln \left(\frac{p_p}{p_n} \right) \end{aligned} \quad (6)$$

Dabei sind V_p und p_p das Potential bzw. die Löcherdichte weit links von der Grenzschicht im p-Bereich, V_n und p_n die Werte weit rechts im n-Bereich. $V_{bi} := V_n - V_p$ ist die so genannte *Diffusionsspannung*, die sich zwischen den zwei Seiten des p-n-übergangs einstellt. Die Bezeichnung V_{bi} kommt vom englischen Ausdruck 'built-in Potential'. Die Löcherdichte weit links im p-Bereich kommt fast ausschließlich von der p-Dotierung. Daher ist in sehr guter Näherung

Diffusionsspannung

$$p_p = N_A. \quad (7)$$

Im rechten n-Bereich ist die Elektronendichte $n_n = N_D$ bekannt. Sie kann mit Hilfe des (nichttrivialen!) Massenwirkungsgesetzes $n_n p_n = n_i^2$ in die Löcherdichte umgerechnet werden:

Massenwirkungsgesetz

$$p_n = \frac{n_i^2}{N_D} \quad (8)$$

Hier ist n_i die 'intrinsische' Elektronen- bzw. Löcherdichte von reinem Silizium von $1.45 \times 10^{10} \text{ cm}^{-3}$ bei Raumtemperatur (300K), die durch thermische Anregung von z.B. Elektronen ins Leitungsband entsteht. Mit (7) und (8) erhält man aus (6) den Wert der Diffusionsspannung als Funktion der Dotierungen

$$V_{bi} = \frac{kT}{q} \ln \left(\frac{N_A N_D}{n_i^2} \right) \quad (9)$$

$$\approx \left[\log \left(\frac{N_A}{n_i} \right) + \log \left(\frac{N_D}{n_i} \right) \right] \times 60 \text{ mV}. \quad (10)$$

Der rechte Ausdruck gilt bei Raumtemperatur. Wie zu erwarten war ist dieser Ausdruck symmetrisch in N_D und N_A . V_{bi} ist immer positiv, die n-Seite des Übergangs liegt also auf positiverem Potential als die p-Seite.

n-Seite positiver als p-Seite

1.3 Betrachtung im Bändermodell mit Hilfe der Fermi-Levels

Die Quantenmechanik teilt die unterschiedlichen Elementarteilchen in zwei fundamentale Kategorien ein: Bosonen und Fermionen. Photonen zum Beispiel sind Bosonen und kommen daher 'gerne' in gleichartigen (Energie- oder Impuls-) Zuständen vor. Das nutzt z.B. ein Laser aus. Elektronen dagegen sind Fermionen. Zwei Fermionen können sich *nicht* im selben Quantenzustand befinden. Bringt man viele Fermionen in ein begrenztes Raumvolumen, so müssen die neu hinzukommenden immer höhere Energieniveaus einnehmen, da

Fermionen und Bosonen

die Niveaus mit niedriger Energie schon besetzt sind. Dies ist z.B. der Grund dafür, daß sich in Atomen die 'Schalen' ausbilden. (Da Elektronen einen Spin von 1/2 haben, kann jedes Energieniveau mit zwei Teilchen mit entgegengesetzten Spins besetzt werden.) Die Wahrscheinlichkeit, daß ein Fermion in einem Ensemble die Energie E hat, wird durch die quantenmechanische *Fermi-Dirac-Verteilung* (Abb. 2) gegeben:

Fermi-Verteilung

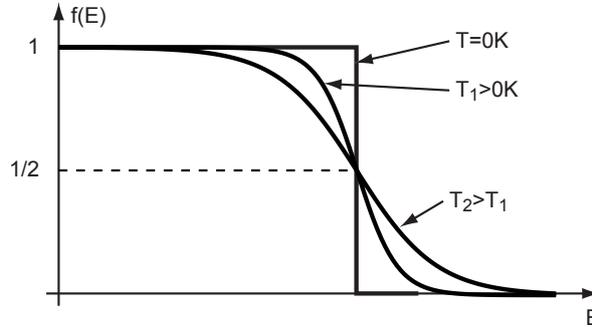


Abbildung 2: Fermi-Dirac-Verteilung für $T_2 > T_1 > T_0 = 0K$

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}} \quad (11)$$

Hierin ist k die Boltzmann-Konstante, T die absolute Temperatur (in Grad Kelvin) und E_F die *Fermi-Energie*. Im Grenzfall $T \rightarrow 0K$ sind alle Niveaus bis zur Fermi-Energie besetzt, die Verteilung wird zu einer Kastenfunktion. Bei höheren Temperaturen schmilzt die 'Fermi-Kante' ab und es gibt Fermionen mit höheren Energien. Die Wahrscheinlichkeit, ein Fermion mit einer Energie von E_F anzutreffen, ist immer 1/2. Für $E - E_F \gg kT$ (also für hohe Fermion-Energien) ist die Exponentialfunktion im Nenner wesentlich grösser als die Eins, so daß diese vernachlässigt werden kann. Die Fermi-Verteilung kann dann durch eine Boltzmann-Verteilung angenähert werden:

Fermi-Energie

$$f(E) \approx e^{-(E-E_F)/kT} \quad \text{mit} \quad E \gg E_F + kT \quad (12)$$

Die Fermi-Verteilung bleibt auch gültig, wenn gewisse Energien nicht besetzt werden

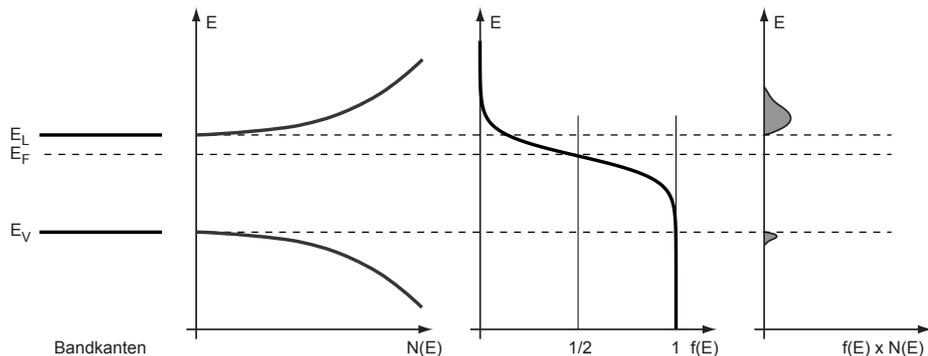


Abbildung 3: Fermiverteilung in der Bandlücke eines n-Halbleiters

können, wie dies z.B. in der Bandlücke eines Halbleiters der Fall ist. Aus Abb. 3 wird ersichtlich, daß z.B. die Anzahl der Elektronen im Leitungsband mit der Fläche unterhalb der Fermiverteilung ab $E = E_L$ zusammenhängt. Da in einem *intrinsischen Halbleiter* die Anzahl Elektronen im Leitungsband gleich der Anzahl Löcher im Valenzband ist, muß die Fermi-Energie in diesem Fall (etwa) in der Mitte der Bandlücke liegen (denn die Fermi-Verteilung ist punktsymmetrisch um $(E_F, 1/2)$). Für einen intrinsischen Halbleiter gilt daher

intrinsischer Halbleiter

$$E_{F,i} \approx \frac{E_L + E_V}{2} \quad (13)$$

Später wird klar werden, daß das nicht ganz exakt gilt. In einem n-dotierten Halbleiter gibt es viel mehr Elektronen im Leitungsband als Löcher im Valenzband. Wir erwarten daher, daß das Fermi-Niveau höher liegt. Um den allgemeinen Fall zu berechnen, muß man beachten, daß aus quantenmechanischen Gründen im Kristallgitter die Dichte von besetzbaren Zuständen energieabhängig ist. Die *Zustandsdichte* ist im Leitungsband (ohne Herleitung)

$$N(E) = \frac{(2m^*)^{3/2}}{2\pi^2} \left(\frac{2\pi}{h}\right)^3 (E - E_L)^{1/2} = C_1 (E - E_L)^{1/2}. \quad (14)$$

m^* ist die *effektive Masse* der Ladungsträger, die sich durch quantenmechanische Effekte von der freien Masse der Ladungsträger unterscheiden kann und die i.A. für Elektronen und Löcher unterschiedlich ist. Die Naturkonstante $h = 6.63 \times 10^{-34} \text{ Js}$ in (14) ist das 'Planck'sche Wirkungsquantum'. Die Anzahl Elektronen im Leitungsband ist also

$$n = \int_{E_L}^{\infty} N(E) f(E) dE \quad (15)$$

$$\approx C_1 \int_{E_L}^{\infty} (E - E_L)^{1/2} e^{-(E - E_F)/kT} dE \quad (16)$$

$$= C_1 e^{-(E_L - E_F)/kT} \int_0^{\infty} y^{1/2} e^{-y/kT} dy$$

$$= C_1 (kT)^{3/2} e^{(E_F - E_L)/kT} \underbrace{\int_0^{\infty} x^{1/2} e^{-x} dx}_{\frac{\sqrt{\pi}}{2}} \quad (17)$$

$$= C_1 (kT)^{3/2} e^{(E_F - E_L)/kT} \cdot \frac{\sqrt{\pi}}{2}$$

$$= \underbrace{2 \left(\frac{2\pi m^* kT}{h^2}\right)^{3/2}}_{N_C} e^{(E_F - E_L)/kT}$$

$$n = N_C \cdot e^{(E_F - E_L)/kT} \quad (18)$$

Dabei haben wir von (15)→(16) die Boltzmann-Näherung (12) benutzt. Nach zweifacher Substitution ($E - E_L = y$, $y/kT = x$) kann das Integral in (17) gelöst werden. Aus (18) bekommt man nun die Lage des Fermi-Niveaus, wenn man annimmt, daß die Elektronen im Leitungsband alle von der Dotierung herrühren, also daß $n = N_D$ ist.

$$E_{F,n} = E_L - kT \ln \frac{N_{C,n}}{N_D} \quad (19)$$

$$E_{F,p} = E_V + kT \ln \frac{N_{C,p}}{N_A} \quad (20)$$

Gleichung 20 ist der entsprechende Ausdruck für das Fermi-Niveau in einem p-dotierten Halbleiter. Die beiden Konstanten $N_{C,n}$ und $N_{C,p}$ müssen unterschieden werden, denn sie enthalten die unterschiedlichen effektiven Massen von Elektronen und Löchern. Sie sind i.A. grösser als N_D oder N_A . Für einen intrinsischen Halbleiter ist $N_D = N_A = n_i$, also gilt für Elektronen und Löcher

$$E_{F,i} = E_L - kT \ln \frac{N_{C,n}}{n_i} \quad (21)$$

$$E_{F,i} = E_V + kT \ln \frac{N_{C,p}}{n_i}. \quad (22)$$

Wenn $N_{C,n} = N_{C,p}$ wäre würde hieraus (13) folgen. Die Abweichung davon aufgrund der unterschiedlichen effektiven Massen von Elektronen und Löchern ist in Silizium klein. Aus den beiden obigen Gleichungen erhält man einen Zusammenhang zwischen der Bandlücke und n_i , $N_{C,n}$ und $N_{C,p}$:

$$E_L - E_V = kT \ln \frac{N_{C,n} N_{C,p}}{n_i^2}. \quad (23)$$

Bringt man p- und n-dotiertes Material in Kontakt, so muß das Fermi-Niveau auf beiden Seiten gleich sein, denn die zugehörige Fermi-Verteilung gilt universell. Da E_F im

Zustandsdichte

effektive Masse

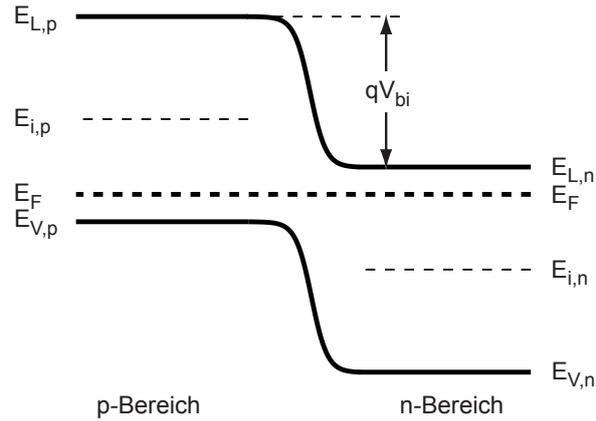


Abbildung 4: Lage der Bänder und des Fermi-Niveaus beim p-n-übergang

n-Bereich kurz unter der Leitungsbandkante liegt, werden die Energieniveaus dort abgesenkt, im p-Bereich werden sie entsprechend angehoben. Dies ist in Abb. 4 skizziert. Der sich ausbildende Potentialunterschied ergibt sich aus der Abbildung zu

$$\begin{aligned} qV_{bi} &= E_{L,p} - E_{L,n} \\ &= \underbrace{E_{L,p} - E_{V,p}} + \underbrace{E_{V,p} - E_F} + \underbrace{E_F - E_{L,n}} \end{aligned} \quad (24)$$

$$= E_L - E_V - kT \ln \frac{N_{C,p}}{N_A} - kT \ln \frac{N_{C,n}}{N_D} \quad (25)$$

$$= kT \ln \frac{N_{C,n} N_{C,p}}{n_i^2} - kT \ln \frac{N_{C,p}}{N_A} - kT \ln \frac{N_{C,n}}{N_D} \quad (26)$$

$$V_{bi} = \frac{kT}{q} \ln \frac{N_A N_D}{n_i n_i} \quad (27)$$

Im Schritt von (24) \rightarrow (25) wurden (19) und (20) benutzt, in (25) \rightarrow (26) wurde (23) eingesetzt und die Logarithmen dann zusammengefasst. Wir erhalten also das gleiche Ergebnis wie in (9).

1.4 Berechnung der Dicke der p-n-Sperrschicht

Die Raumladungszone endet in Wirklichkeit nicht abrupt, sondern innerhalb einer dünnen Übergangszone. Diese ist jedoch so dünn, dass die Annahme des abrupten Übergangs sehr gut gerechtfertigt ist. Die (positive) Gesamtladung im n-Bereich der Raumladungszone (mit der Dicke x_n) beträgt bei einer Diode der Fläche A

$$Q_n = Ax_n q N_D > 0. \quad (28)$$

Für den p-Bereich gilt wegen $x_p < 0$ vorzeichenrichtig

$$Q_p = Ax_p q N_A < 0. \quad (29)$$

Die beiden Raumladungszonen entstehen durch das Wegwandern von Elektronen bzw. Löchern. Da diese wegen Ladungserhaltung nicht verloren gehen können, muß $Q_p = -Q_n$ sein, also

$$\begin{aligned} Q_p &= -Q_n \\ Ax_p q N_A &= -Ax_n q N_D \\ x_p N_A &= -x_n N_D. \end{aligned} \quad (30)$$

Durch die Raumladungszonen entsteht ein elektrisches Feld, das an der Grenzfläche seinen höchsten Wert E_{max} hat. Dieser kann mit Hilfe des Gauß'schen Satzes berechnet werden. Dazu betrachten wir eine Dose der Fläche A , die z.B. weit links im (nicht verarmten)

p-Bereich beginnt und an der Grenzschicht endet (s. Abb. 1). Das Integral über die austretenden Felder enthält nur den Beitrag $A \cdot E_{max}$ der rechten Dosenfläche (da links das Feld Null ist und am Rand der Dose die Feldlinien parallel zur Oberfläche verlaufen). Die eingeschlossene Ladung ist durch (29) gegeben. Der Gauß'sche Satz liefert

$$A \cdot E_{max} = \frac{Ax_p q N_A}{\epsilon} \Rightarrow E_{max} = \frac{q}{\epsilon} x_p N_A < 0. \quad (31)$$

Hierbei ist $\epsilon = \epsilon_S \epsilon_0$. Die Feldstärke ist negativ, die Feldlinien zeigen also, wie in der Einleitung bereits vermutet, nach links. Wie in Abb. 1 angedeutet ist, fällt das Feld also von x_p an von Null linear auf E_{max} ab und steigt dann wieder linear an, bis es bei x_n wieder Null ist. Der *Potentialunterschied* zwischen der linken und der rechten Seite lässt sich nun durch Integration des Feldes berechnen. Das Integral ist einfach die Fläche unter der dreieckigen $E(x)$ -Kurve, also

$$\Delta V = V_n - V_p = - \int_{x_p}^{x_n} E(x) dx = -\frac{1}{2} (x_n - x_p) E_{max} > 0. \quad (32)$$

Die Spannung ist also rechts positiver, wie es aufgrund der positiven Raumladung im n-Bereich ja auch sein muß. Nun kann man in (32) z.B. x_n mit Hilfe von (30) durch x_p ausdrücken. Nach Einsetzen von (31) erhält man

$$\Delta V = \frac{1}{2} \left(\frac{x_p N_A}{N_D} + x_p \right) \frac{q}{\epsilon} x_p N_A = \frac{q}{2\epsilon} \frac{(N_A + N_D) N_A}{N_D} x_p^2 \quad (33)$$

oder

$$|x_p| = \sqrt{\frac{2\epsilon}{q} \frac{N_D}{(N_A + N_D) N_A} \Delta V} \xrightarrow{N_D \rightarrow \infty} \sqrt{\frac{2\epsilon}{q} \frac{\Delta V}{N_A}} \quad (34)$$

$$|x_n| = \sqrt{\frac{2\epsilon}{q} \frac{N_A}{(N_A + N_D) N_D} \Delta V} \xrightarrow{N_A \rightarrow \infty} \sqrt{\frac{2\epsilon}{q} \frac{\Delta V}{N_D}}. \quad (35)$$

Die Verarmungszonen sind also um so dünner, je stärker dotiert die Bereiche sind (das ist ja bereits aus (30) klar). Die Gesamtdicke x_d der Sperrschicht ergibt sich dann zu

Dicke der Sperrschicht

$$\begin{aligned} x_d &= x_n - x_p \stackrel{(32)}{=} - \frac{2\Delta V}{E_{max}} \stackrel{(31)}{=} - \frac{2\epsilon\Delta V}{q x_p N_A} \stackrel{(33)}{=} \frac{2\epsilon\Delta V}{q N_A} \sqrt{\frac{q}{2\epsilon\Delta V} \frac{(N_A + N_D) N_A}{N_D}} \\ &= \sqrt{\frac{2\epsilon}{q} \frac{N_A + N_D}{N_A N_D} \Delta V}. \end{aligned} \quad (36)$$

Die Spannung ΔV setzt sich aus der Diffusionsspannung V_{bi} und einer von außen angelegten externen Spannung V_{ext} zusammen:

Vorzeichendefinition

$$\Delta V = V_{bi} - V_{ext}. \quad (37)$$

Dabei haben wir das Vorzeichen so gewählt, daß ΔV für $V_{ext} = V_{bi}$ Null wird. Der positive Pol der externen Spannung liegt also an der p-Seite, der *Anode*, an. Mit einer externen Spannung kann also die Dicke der Sperrschicht verändert werden:

$$\begin{aligned} x_d &= \underbrace{\sqrt{\frac{2\epsilon}{q} \frac{N_A + N_D}{N_A N_D} V_{bi}}}_{x_d(0)} \sqrt{1 - \frac{V_{ext}}{V_{bi}}} \\ &= x_d(0) \left(1 - \frac{V_{ext}}{V_{bi}} \right)^{1/2} \end{aligned} \quad (38)$$

Für $V_{ext} = V_{bi}$ verschwindet die Sperrschicht ganz.

1.5 Die Kapazität der Sperrschicht

Wir berechnen hier nur die Kapazität einer in Sperrrichtung gepolten Diode, bei der sich eine von beweglichen Ladungsträgern freie Raumladungszone ausgebildet hat und in der daher (außer einem kleinen Leckstrom) kein Strom fließt. Im VLSI Design sind Dioden fast immer in Sperrrichtung geschaltet und bilden eigentlich unerwünschte 'parasitäre' Elemente, die insbesondere durch ihre Kapazität das Verhalten der Schaltung (meist negativ) beeinflussen.

Die Diode stellt einen Plattenkondensator mit dem Plattenabstand x_d dar, der mit Silizium mit der Dielektrizitätskonstante ϵ_{Si} gefüllt ist. Die Kapazität ist daher

$$C_j = \epsilon_0 \cdot \epsilon_{Si} \cdot \frac{A}{x_d} \quad (39)$$

$$\begin{aligned} \frac{C_j}{A} &= \sqrt{\frac{q\epsilon_0\epsilon_{Si}}{2} \frac{N_A N_D}{N_A + N_D} \frac{1}{V_{bi}}} \left(1 - \frac{V_{ext}}{V_{bi}}\right)^{-1/2} \\ &= C_{j0} \cdot \left(1 - \frac{V_{ext}}{V_{bi}}\right)^{-1/2} \end{aligned} \quad (40)$$

Der Index 'j' zeigt an, dass es sich um die 'junction'-Kapazität handelt. C_{j0} hat die Einheit F/m^2 . Dieses Ergebnis wurde für einen abrupten Übergang vom p- ins n-dotierte Gebiet hergeleitet. In der Praxis ist der Verlauf fließend (je nach Herstellungsverfahren) wodurch insbesondere der Exponent in (40) nicht mehr 1/2, sondern kleiner ist. Aus diesem Ergebnis wird sichtbar, dass die Kapazität mit steigenden Dotierungen sinkt. Schwache Dotierungen sind hier also von Vorteil!

Aufgrund dieser Ergebnisse wird in SPICE die Kapazität einer p-n-Diode mit folgender Formel berechnet:

$$C = A \times \mathbf{CJ} \times \left(1 - \frac{V}{\mathbf{PHA}}\right)^{-\mathbf{MJ}} + P \times \mathbf{CJSW} \times \left(1 - \frac{V}{\mathbf{PHP}}\right)^{-\mathbf{MJSW}} \quad (41)$$

Hierin ist V die angelegte Spannung. Die SPICE Modellparameter **CJ** und **MJ** und das (Diffusions-) Potential **PHA** gelten für den 'Boden' der Diode. Dieser Anteil skaliert also mit der Fläche A (Area) der Diode. Am Rand der Diode P (periphery) sind die Dotierungsverhältnisse meist anders, so dass hierfür unabhängige Parameter **CJSW**, **PHP** und **MJSW** vorhanden sind. Bei kleinen Dioden sind beide Beträge oft vergleichbar!

1.6 Zahlenwerte

Wir betrachten einen abrupten p-n-Übergang mit einer (schwachen) n-Dotierung von $N_D = 10^{16} \text{ cm}^{-3}$ und einer (starken) p-Dotierung von $N_A = 10^{19} \text{ cm}^{-3}$. Die Diffusionsspannung ist nach (10)

$$\begin{aligned} V_{bi} &= \left[\log\left(\frac{10^{19}}{1.45 \times 10^{10}}\right) + \log\left(\frac{10^{16}}{1.45 \times 10^{10}}\right) \right] \times 60 \text{ mV} \\ &= [8.84 + 5.84] \times 60 \text{ mV} = 0.874 \text{ V}. \end{aligned} \quad (42)$$

Die Verarmungszone erstreckt sich fast nur in den schwächer dotierten n-Bereich: Nach (30) ist

$$\frac{|x_p|}{|x_n|} = \frac{N_D}{N_A} = 10^{-3},$$

der p-dotierte Bereich ist also 1000 mal dünner! Gleichung 35 liefert mit $\epsilon = \epsilon_{Si}\epsilon_0 = 11.9 \times 8.854 \times 10^{-12} \text{ F/m} = 1.05 \times 10^{-10} \text{ F/m}$ ohne äussere Spannung, also für $\Delta V = V_{bi}$

$$\begin{aligned} x_d(0V) \approx x_n(0V) &\approx \sqrt{\frac{2\epsilon V_{bi}}{q N_D}} \\ &= \sqrt{\frac{2 \times 1.05 \times 10^{-10} \text{ F/m} \times 0.874 \text{ V}}{1.6 \times 10^{-19} \text{ C} \times 10^{22} \text{ m}^{-3}}} \\ &= 0.34 \mu\text{m}. \end{aligned} \quad (43)$$

Die Dotierungsdichte muss dabei von der gebräuchlichen Einheit cm^{-3} auf m^{-3} umgerechnet werden! Bei einer extern angelegten Sperrspannung von $V_{\text{ext}} = -2\text{V}$ wächst die Dicke der Sperrschicht nach (38) auf

$$x_d(-2\text{V}) \approx x_d(0\text{V}) \times \sqrt{1 - \frac{V_{\text{ext}}}{V_{bi}}} = 0.61\ \mu\text{m}. \quad (44)$$

Die maximale Feldstärke (an der Grenzfläche) beträgt nach (31)

$$|E_{\text{max}}| = \frac{q}{\epsilon} x_p N_A = \frac{q}{\epsilon} x_n N_D = 5.2 \times 10^6\ \text{V/m} \quad (45)$$

Das liegt noch deutlich unter der Durchbruchfeldstärke von Silizium von $3 \times 10^7\ \text{V/m}$. Die Flächenkapazität einer $2 \times 2\ \mu\text{m}^2$ grossen Diode beträgt nach (39)

$$C_j \approx \frac{\epsilon A}{x_d} = 1.2 \times 10^{-15}\ \text{F} = 1.2\ \text{fF} \quad (46)$$

Der Rand dieser Diode ist $4 \times 2\ \mu\text{m}$ lang und hat mit seiner Tiefe von $0.34\ \mu\text{m}$ somit eine zusätzliche Fläche von $2.7\ \mu\text{m}^2$. Bei gleichen Dotierungsprofilen erhöht der Rand also die Kapazität um zusätzlich 70%.

1.7 Herleitung der Kennlinie der Diode

Im VLSI Design sind Dioden fast immer in Sperrrichtung geschaltet. Wichtig sind für uns daher vor allem die Kapazitätseigenschaften. Trotzdem soll hier die Kennlinie der Diode, also der Strom als Funktion der extern angelegten Spannung, hergeleitet werden. Ein solcher Strom kommt zustande, wenn das Gleichgewicht zwischen Diffusions- und Feldstrom durch eine externe Spannung gestört wird. Wie in Abb. 5 skizziert ist (und in den vorherigen Abschnitten berechnet wurde) wird die Sperrschicht durch eine extern angelegte Spannung dicker, wenn eine positive externe 'Rückwärts'-Spannung (Index 'R' für 'Reverse') an die n-Seite angelegt wird. Die Sperrschicht wird *dünnere*, wenn eine *positive* 'Vorwärts'-Spannung (Index 'F' für 'Forward') an die *p-Seite* angelegt wird. In diesem Fall erhöht sich die Wahrscheinlichkeit, daß Ladungsträger durch Diffusion die (dünnere) Barriere überwinden, der Diffusionsstrom steigt also an. Der Feldstrom dagegen wird bei der Diode im Wesentlichen durch die Zahl der zur Verfügung stehenden Ladungsträger bestimmt (das wurde bisher nicht so klar und ist nicht selbstverständlich!). Er bleibt daher etwa konstant. Insgesamt ergibt sich also ein Vorwärtsstrom, der um so grösser ist, je höher die angelegte Vorwärtsspannung ist.

Vorwärtsspannung
Diffusionsstrom steigt
Feldstrom bleibt

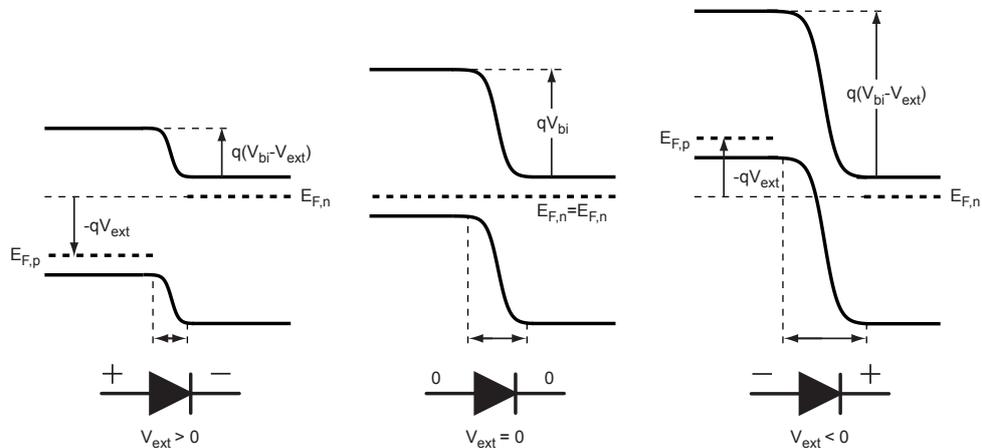


Abbildung 5: Lage der Bänder und Dicke der Grenzschicht bei angelegter Sperrspannung (rechts) und Vorwärtsspannung (links)

Für eine quantitative Herleitung nutzen wir aus, daß es einen festen Zusammenhang zwischen der Spannung über der Verarmungszone und Elektronen- und Löcherdichten gibt. Um diese Bedingung auch bei angelegter externer Spannung zu erfüllen, müssen sich die Ladungsträgerkonzentrationen verändern. Dies führt zu einer erhöhten Elektronendichte im p-Bereich nahe der Grenzschicht und zu einer erhöhten Löcherdichte im n-Bereich. Diese 'überschüssigen' Ladungsträger können in die Sperrschicht diffundieren und werden dann aufgrund des dort vorhandenen Feldes zur anderen Seite gesaugt. So entsteht ein Strom durch die Sperrschicht. Er hängt von der Dichte der 'überschüssigen' Ladungsträger ab, welche wiederum von der extern angelegten Spannung abhängt.

Zur Herleitung bezeichnen wir Elektronendichten mit n_{index} , Löcherdichten mit p_{index} . Der Index gibt an, an welcher Stelle die Dichte vorliegt, so ist z.B. p_{po} die Löcherdichte weit im p-Bereich, also weit links. Es gilt die Geometrie von Abb. 1. Ausgangspunkt unserer Überlegungen ist die wichtige Gleichung (6), die durch Integration von einer Stelle weit links von der Grenzschicht bis zu einer Stelle weit rechts erhalten wurde. Mit der hier definierten Schreibweise lautet (6) also

$$V_{bi} = V_n - V_p = \frac{kT}{q} \ln \left(\frac{p_{po}}{p_{no}} \right)$$

oder

$$p_{po} = p_{no} e^{\frac{qV_{bi}}{kT}} \tag{47}$$

$$n_{po} = n_{no} e^{-\frac{qV_{bi}}{kT}} \tag{48}$$

Gleichung (48) für die Elektronendichten erhält man unter Verwendung des Massenwirkungsgesetzes. Diese beiden Gleichungen sagen aus, daß zu einem Spannungsunterschied immer ein Unterschied in den Konzentrationen der Ladungsträger gehört. Die Löcherdichte z.B. ist nach (47) im p-Bereich wesentlich grösser als im n-Bereich ($V_{bi} > 0!$) wie es sein muß. Legt man eine externe Vorwärtsspannung $V_{ext} = V_F > 0$ an, so verringert sich das eingebaute Potential auf $V_{bi} - V_{ext}$. Wir nehmen an, daß diese Spannung komplett über der Sperrschicht abfällt. Damit die obigen Gleichungen weiterhin erfüllt sind, muß sich z.B. die Elektronenkonzentration so verändern, daß am linken und rechten Rand der Grenzschicht gilt:

$$n_n = n_p e^{\frac{q(V_{bi}-V_{ext})}{kT}} \quad (49)$$

Darin sind n_n und n_p die Elektronendichten kurz ausserhalb der Grenzschicht im n- und p-Bereich.

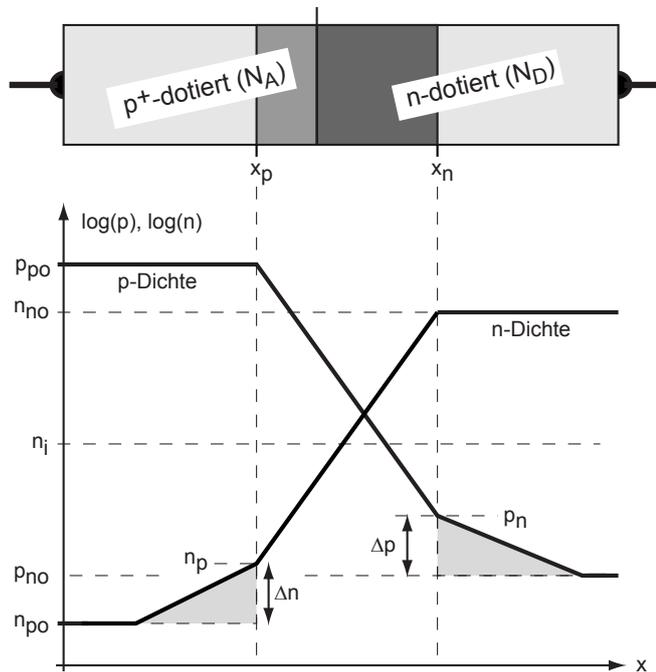


Abbildung 6: Ladungsträgerdichten im p-n-übergang bei Anlegen einer Spannung in Vorwärts-Richtung

Die Elektronendichte ist im n-Bereich so hoch, daßs sie durch das Abwandern einiger weniger Elektronen praktisch unverändert bleibt, also $n_n \approx n_{no}$ ist. Unter Benutzung von (48) wird daher aus (49)

$$\begin{aligned} n_{no} &\approx n_n \\ n_{po} e^{\frac{qV_{bi}}{kT}} &\approx n_p e^{\frac{q(V_{bi}-V_{ext})}{kT}} \\ n_p &\approx n_{po} e^{\frac{qV_{ext}}{kT}}. \end{aligned} \quad (50)$$

Die Dichte der *Elektronen* an der *linken p-Seite* der Grenzschicht n_p (Elektronen sind im dortigen p-Material Minoritätsträger!) ist also durch die positive externe Vorwärtsspannung *höher* als weit links im Material. Der Zugewinn im Vergleich zum Gleichgewichtszustand mit $V_{ext} = 0$ V beträgt

$$\Delta n = n_p - n_{po} = n_{po}(e^{\frac{qV_{ext}}{kT}} - 1). \quad (51)$$

überschuß an Elektronen
im p-Bereich

Dies ist in Abb. 6 eingezeichnet. Entsprechendes gilt für die Löcherdichte an der n-Seite. Die an der Grenzschicht erhöhte Dichte an Minoritätsträgern sinkt im nicht-verarmten

Injektion von
Minoritätsträgern

Bereich schnell auf den Gleichgewichtswert ab, wie das in Abb. 6 angedeutet ist. Zur Berechnung muß man dazu eine Diffusionsgleichung lösen. Man findet einen exponentiellen Abfall, da sich die Minoritätsträger durch Rekombination nicht weit bewegen können. Die Minoritätsträger diffundieren auch in die Sperrschicht, in der ja ein elektrisches Feld herrscht. Elektronen im p-Bereich z.B. werden dann zum positiven n-Bereich hin abgesaugt. Man sagt, daß Ladungsträger in die Sperrschicht *injiziert* werden. Beide Sorten von Ladungsträgern tragen also zu einem Strom bei, der proportional zum überschüss (51) an Ladungsträgern sein muß. Also gilt

$$I_D = I_0 (e^{\frac{qV}{kT}} - 1). \quad (52)$$

Dies ist die bekannte Diodengleichung. (Bei der genaueren Herleitung muß man zunächst die Dichten der Ladungsträger weiter weg von der Grenzschicht durch Lösen der Diffusionsgleichung bestimmen und daraus den Diffusionsstrom an der Grenzschicht bestimmen.) Für genügend hohe Spannungen kann man die Eins in (52) vernachlässigen. Man hat dann einen exponentiellen Anstieg des Diodenstroms mit der angelegten Vorwärtsspannung. Der Strom verzehnfacht sich, wenn die Diodenspannung um 60 mV erhöht wird.

1.8 Zusammenfassung

Hier sind die wichtigsten Ergebnisse zusammengefasst:

- Die Diffusionsspannung, die sich an der pn-Grenzschicht ausbildet, hängt nur von den Dotierungen ab:

$$V_{bi} = \frac{kT}{q} \ln \left(\frac{N_A N_D}{n_i^2} \right) = \frac{kT}{q} \ln \left(\frac{N_A}{n_i} \right) + \frac{kT}{q} \ln \left(\frac{N_D}{n_i} \right)$$

- Die Diode leitet, wenn eine positive Spannung am p-Bereich anliegt
- Der Vorwärtsstrom steigt exponentiell an nach der Formel

$$I = I_S (e^{V/V_{th}} - 1)$$

- $V_{th} = kT/q \approx 26 \text{ mV}$ bei 300K. Der Strom verzehnfacht sich also etwa alle 60 mV
- Das E-Feld ist am Übergang am höchsten
- Die Sperrschicht wächst in den schwach dotierten Bereich. Sie ist um so dicker, je schwächer die Dotierung ist. Die Dicke der Sperrschicht im schwach dotierten Bereich ist

$$d \approx \sqrt{\frac{2\epsilon V_{bi} + V_{ext}}{q N_{D/A}}}$$

- Die Sperrschicht wächst wie die Wurzel aus der Spannung (mit einem Offset durch die Diffusionsspannung)
- Die Kapazität der Sperrschicht sinkt mit der Wurzel aus der angelegten Spannung
- Bei integrierten Dioden berücksichtigt man einen Beitrag durch die Fläche und einen durch den Rand

2 Der Sperrschicht-Feldeffekttransistor (JFET)

In einem Sperrschicht-Feldeffekttransistor (Junction-FET, JFET) wird der Querschnitt eines leitenden Kanals durch eine von der Seite eindringende Raumladungszone verändert. Die Raumladungszone ist die Sperrschicht einer p-n-Diode. Sie wird über eine an die Gate-Elektrode angelegte Sperrspannung verändert. In Abb. 7 ist der Aufbau eines n-Kanal JFET schematisch dargestellt. Ein Stück schwach n-dotiertes Silizium der Dicke

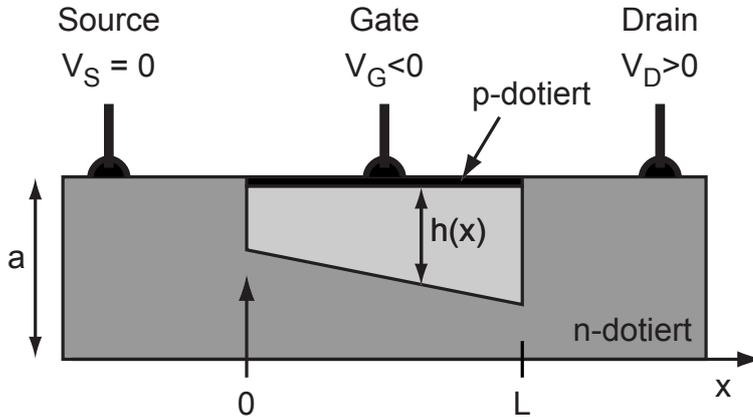


Abbildung 7: Stark schematisierter Aufbau eines n-Kanal JFET. Der verarmte Bereich ist hellgrau gezeichnet.

a ist an *Source* und *Drain* kontaktiert, so dass zwischen diesen Elektroden bei einer angelegten Spannung ein Strom durch den *Kanal* fließen kann. Wir vereinbaren, dass die Source auf Massepotential liegt ($V_S = 0\text{ V}$). An der Oberfläche des Siliziums ist eine dünne, stark p-dotierte Schicht der Länge L eingebaut, die mit der *Gate* Elektrode verbunden ist. An der Grenzfläche zum Kanal bildet sich eine ladungsträgerfreie und daher isolierende Verarmungszone aus, deren Dicke vom Potentialunterschied zwischen dem Gate und dem *Potential im Kanal* abhängt. Für $V_G = 0$ ist dieser Potentialunterschied Null an der Source-Seite. Er steigt auf $V_D > 0$ an der Drain-Seite an, so dass die Raumladungszone dort dicker ist (s. Abb. 7). Mit einer negativen Spannung am Gate (bei stark positiver Spannung würde die Gate-Kanal-Diode leiten!) wächst die Verarmungszone in den n-Bereich und verkleinert dort den Querschnitt des Kanals, so dass der Drainstrom abnimmt. Ab einer bestimmten Gatespannung (der 'Pinch-Off'-Spannung V_p) versperrt die Raumladungszone den gesamten Kanal und der Stromfluss wird vollständig unterbrochen. Das Abschnüren des Kanals bei einer stark negativen Gatespannung ist in der Übertragungskennlinie in Abb. 8 links schematisch dargestellt. Bei einer Gatespannung $V_G > V_p$ ist der Kanal an der Source-Seite vorhanden. An der Drain-Seite hängt die Dicke der Raumladungszone und damit der Querschnitt des Kanals aber auch von der Drainspannung ab. Für $V_D = V_G - V_p$ schnürt der Kanal lokal ab. Eine weitere Erhöhung von V_D führt nicht zu einer Erhöhung des Potentials am drainseitigen Ende des Kanals (links vom Abschnürpunkt), so dass der Strom nicht mehr ansteigt. Dies ist in der Ausgangskennlinie in Abb. 8 rechts dargestellt. Den Bereich (fast) konstanten Drainstroms oberhalb von $V_D = V_G - V_p$ nennt man den *Sättigungsbereich*, den Bereich in der Nähe von $V_D = 0$, wo der Drainstrom fast linear mit der Drainspannung zusammenhängt, den *linearen Bereich*.

Source, Drain
Kanal
Gate

Kanalpotential

Abschnüren, Pinch-Off

Sättigungsbereich

Linearer Bereich

2.1 Berechnung der Kennlinie

Zur näherungsweisen Berechnung der Kennlinie betrachten wir nur den Bereich unter der Gate-Dotierung, wir vernachlässigen also den (nahezu konstanten) Zuleitungswiderstand von den Drain/Source-Kontakten bis zum Kanal. Die untere Begrenzung des Kanals wird in der praktischen Realisierung des JFET durch eine weitere, großflächig p-dotierte Schicht erreicht, an der sich eine weitere Raumladungszone ausbildet. Die Ergebnisse dieses Abschnitts ändern sich dadurch ein wenig.

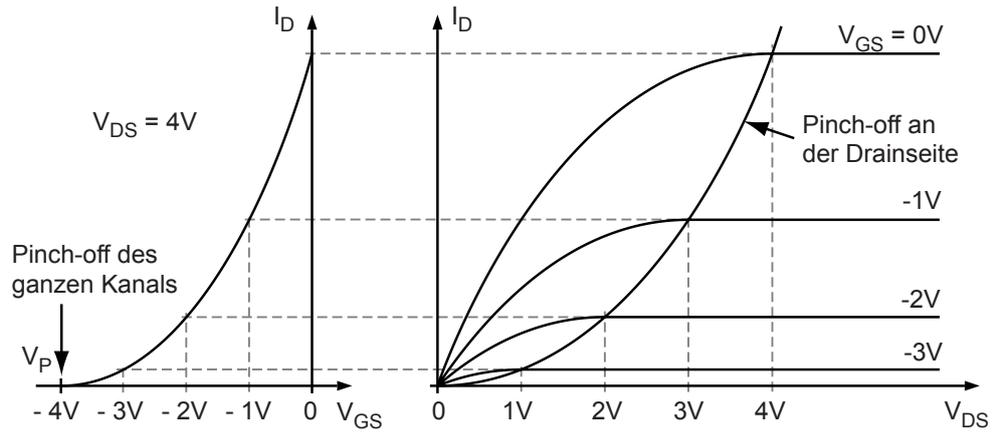


Abbildung 8: Übertragungskennlinie (links) und Ausgangskennlinie (rechts) eines n-Kanal JFET

Wir halten zunächst noch einmal fest, dass der n-Kanal JFET nur mit negativen Gate-Spannungen betrieben werden kann, da nur so die Gate-Kanal-Diode sperrt:

$$V_G < 0 \quad (53)$$

Kanalpotential Das Potential im Kanal $V(x)$ ist Null bei der Source (bei $x = 0$) und steigt auf $V_D > 0$ bei $x = L$ an. Es ist also nicht konstant. Die an der Gate-Kanal-Diode anliegende Spannung ist somit $V_{ext} = V(x) - V_G$. Die Raumladungszone hat daher nach (35) an der Position x im Kanal eine Dicke von

$$h(x) = \left[\frac{2\epsilon}{qN_D} (V(x) - V_G + V_{bi}) \right]^{1/2}. \quad (54)$$

Hier ist $V_{bi} > 0$ das Diffusionspotential der Gate-Kanal-Diode. Die Spannungen gehen mit positivem Vorzeichen ein, weil das Gate p-dotiert ist und die Diode in Sperrichtung polarisiert ist. Zur einfacheren Schreibweise fassen wir die Gatespannung und das dazu additive Diffusionspotential in $V'_G > 0$ zusammen:

$$V'_G := V_{bi} - V_G > 0 \quad (55)$$

Da das Kanalpotential von links nach rechts ansteigt wird die Verarmungszone nach rechts immer dicker, wie dies in Abb. 7 angedeutet ist. Die Dicken an Source und Drain sind insbesondere

$$h_S = h(0) = \left[\frac{2\epsilon}{qN_D} V'_G \right]^{1/2} \quad (56)$$

$$h_D = h(L) = \left[\frac{2\epsilon}{qN_D} (V'_G + V_D) \right]^{1/2} \quad (57)$$

Bei genügend hoher Spannung füllt die Verarmungszone zunächst an der Drain-Seite die gesamte Dicke a des Kanals aus. Die ist der Fall, wenn

$$h_D = \left[\frac{2\epsilon}{qN_D} (V'_G + V_D) \right]^{1/2} = a$$

$$V'_G + V_D = \frac{qa^2 N_D}{2\epsilon} =: V_p \quad (58)$$

Abschnürspannung Hier haben wir die häufig benutzte *Abschnürspannung* (engl. pinch-off-voltage) V_p eingeführt. (Für $V_D = 0$ ist das die Gatespannung, bei der der gesamte Kanal gleichzeitig abschnürt, für $V_D > 0$ gibt (58) an, wann der Kanal an der Drainseite abschnürt.) Unter Benutzung von V_p können wir die Dicken der Verarmungszonen bei Source (56) und Drain (57) einfacher schreiben:

$$\frac{h_S}{a} = \sqrt{\frac{V'_G}{V_p}} \quad \text{und} \quad \frac{h_D}{a} = \sqrt{\frac{V'_G + V_D}{V_p}} \quad (59)$$

Nach diesen Vorarbeiten betrachten wir den Stromfluß im Kanal. Die horizontal fließende Stromdichte an einer Stelle x im Kanal ist (vgl. (1))

$$j(x) = q N_D \mu E_x(x) = q N_D \mu \frac{dV(x)}{dx} \quad (60)$$

Dabei ist $\mu = \mu_e$ die Mobilität der im Kanal fließenden Elektronen. Der gesamte Strom ergibt sich durch Multiplikation mit der Querschnittsfläche $W(a - h(x))$. Hier ist W die *Breite* (engl. width) des Gates. Der Gesamtstrom muß (Ladungserhaltung!) an jeder Stelle des Kanals gleich sein, er ist also unabhängig von x :

$$I_D = W(a - h(x)) q N_D \mu \frac{dV(x)}{dx} \quad (61)$$

Das Differential $dV(x)$ erhält man, indem man in (54) $dh(x)/dx$ bildet und die Kettenregel verwendet:

$$dV(x) = \frac{qN_D}{\epsilon} h dh \quad (62)$$

Einsetzen von (62) in (61) liefert

$$I_D dx = \frac{Wq^2 N_D^2 \mu}{\epsilon} (a - h) h dh$$

Integration beider Seiten von der Source bis zur Drain liefert

$$\begin{aligned} \int_0^L I_D dx &= \frac{Wq^2 N_D^2 \mu}{\epsilon} \int_{h_s}^{h_D} (a - h) h dh \\ I_D L &= \frac{Wq^2 N_D^2 \mu}{\epsilon} \left[\frac{ah^2}{2} - \frac{h^3}{3} \right]_{h_s}^{h_D} \\ I_D &= \frac{W}{L} \frac{q^2 N_D^2 \mu a^3}{2\epsilon} \left[\frac{h_D^2 - h_S^2}{a^2} - \frac{2}{3} \frac{h_D^3 - h_S^3}{a^3} \right] \end{aligned}$$

Nach Einsetzen der Ausdrücke aus (59) ergibt sich der Drainstrom im JFET im linearen Bereich zu

$$I_D = \frac{W}{L} I_0 \left[\frac{V_D}{V_p} + \frac{2}{3} \left(\frac{V'_G}{V_p} \right)^{3/2} - \frac{2}{3} \left(\frac{V_D + V'_G}{V_p} \right)^{3/2} \right]. \quad (63)$$

2.2 Interpretation des Ergebnisses

Wir wollen dieses Ergebnis nun analysieren:

- Der Strom ist proportional zur W/L , also dem Verhältnis von Breite und Länge des Gates. Die Abhängigkeit von W ist trivial (Vergrößerung von W entspricht die Parallelschaltung mehrerer Transistoren), die Abhängigkeit von L weniger. Der gleiche Zusammenhang wird uns auch beim Feldeffekttransistor begegnen. Die geschickte Wahl dieser Parameter ist beim Entwurf von Schaltungen von zentraler Bedeutung.
- Der Strom hängt sowohl von V_G wie von V_D ab. Die Abhängigkeit von V_D ist nicht einfach linear (wie bei einem einfachen Widerstand), da das Zusammenspiel der beiden Spannungen am drainseitigen Ende den Querschnitt des Kanals festlegt.
- Für $V_D = 0V$ darf natürlich kein Strom fließen. Dies wird von (63) erfüllt, weil sich die beiden rechten Terme wegheben.
- Beim Pinch-Off muß man zwei Fälle unterscheiden: Der Kanal schnürt *vollständig* (also auf der gesamten Länge) ab, wenn die Gate-Kanal-Spannung *überall* die Pinch-Off-Spannung erreicht. Insbesondere muß dies also auch an der Source-Seite der Fall sein, wo die Kanalspannung $0V$ ist. In diesem Fall fließt dann kein Strom mehr, unabhängig von der Drain-Spannung. Im zweiten Fall wird nur an der Drain-Seite die Pinch-Off-Bedingung (58) erfüllt. Es fließt weiterhin ein Strom durch den Kanal, dieser steigt aber bei einer Erhöhung der Drainspannung nicht weiter an (Sättigung). Dies ist in Abb. 8 angedeutet. Die Differenz aus Drain- und Gatespannung bei Erreichen des Pinch-Off ist in diesem Beispiel immer $4V$.

Gatebreite

- Erreicht das Gate die Pinch-Off-Spannung, also $V'_G = V_p$, so erhält man durch Reihenentwicklung

$$\begin{aligned} I_D &= \frac{W}{L} I_0 \left[\frac{V_D}{V_p} + \frac{2}{3} - \frac{2}{3} \left(1 + \frac{V_D}{V_p} \right)^{3/2} \right] \\ &\approx \frac{W}{L} I_0 \left[\frac{V_D}{V_p} + \frac{2}{3} - \frac{2}{3} \left(1 + \frac{3}{2} \frac{V_D}{V_p} + \dots \left(\frac{V_D}{V_p} \right)^2 + \dots \right) \right] \\ &= -\frac{2}{3} \frac{W}{L} I_0 \left(\dots \left(\frac{V_D}{V_p} \right)^2 + \dots \right). \end{aligned}$$

Für positive V_D ist I_D nach diesem Ergebnis negativ, was offensichtlich keinen Sinn macht. Der Ausdruck für den Drainstrom (63) hat hier seinen Gültigkeitsbereich verloren. Der Drainstrom ist daher für alle Drainspannungen Null, wie man es am Pinch-Off Punkt (des gesamten Kanals) auch erwartet.

- Für sehr kleine Drainspannungen, genauer für $V_D \ll V'_G$, kann man den dritten Term in (63) entwickeln:

$$\begin{aligned} \left(\frac{V_D + V'_G}{V_p} \right)^{3/2} &= \left[\frac{V'_G}{V_p} \left(1 + \frac{V_D}{V'_G} \right) \right]^{3/2} \\ &\approx \left(\frac{V'_G}{V_p} \right)^{3/2} \left(1 + \frac{3}{2} \frac{V_D}{V'_G} \right). \end{aligned}$$

Damit erhält man

$$\begin{aligned} I_D &= \frac{W}{L} I_0 \left[\frac{V_D}{V_p} - \left(\frac{V'_G}{V_p} \right)^{3/2} \frac{V_D}{V'_G} \right] \\ &= \frac{W}{L} \frac{I_0}{V_p} \left[1 - \sqrt{\frac{V'_G}{V_p}} \right] V_D \end{aligned} \quad (64)$$

linearer Bereich

Der Drainstrom ist also proportional zur Drainspannung. Der Transistor verhält sich wie ein Widerstand. Die Kennlinie im $I_D(V_D)$ ist in der Nähe des Ursprungs linear. Man nennt diesen Arbeitsbereich den *linearen* oder *ohm'schen Bereich*. Im Englischen spricht man von 'linear' oder 'ohmic region' oder von der *'triode region'*. Auch in Gleichung (64) sieht man übrigens, daß der Strom für $V'_G = V_p$ Null wird.

- Nach (58) schnürt der Kanal an der Drainseite ab, wenn $V_D = V_p - V'_G$. Mit diesen Werten für die Drainspannung wird aus (63)

$$I_D = \frac{W}{L} I_0 \left[\frac{1}{3} + \frac{2}{3} \left(\frac{V'_G}{V_p} \right)^{3/2} - \frac{V'_G}{V_p} \right] \quad (65)$$

Diese Gleichung gibt den Sättigungsstrom als Funktion der Gatespannung an. Wie wir es erwarten fällt der Strom bei sinkender Gatespannung (also steigendem V'_G) ab. Für $V'_G = V_p$ wird er Null: der Transistor ist vollständig abgeschaltet.

Sättigung

- Der Drainstrom steigt mit der Drainspannung an, bis der Kanal am drainseitigen Ende abschnürt, denn dort ist die Spannungsdifferenz zwischen der Drain und dem (negativen) Gate am höchsten. Bei einer weiteren Erhöhung der Drainspannung ändert sich im Kanal praktisch nichts, die zusätzliche Spannung fällt über dem abgeschnürten Bereich ab. Der Strom steigt daher nicht weiter an, der Transistor ist in *Sättigung* (engl.: saturation).

3 Der MOSFET

Das heute wichtigste Bauelement im VLSI Design ist der MOSFET. 'FET' steht für Feld-Effekt-Transistor, was andeutet, daßs der Stromfluß im Kanal über ein extern generiertes elektrisches Feld beeinflusst wird (also stromlos). Das 'MOS' steht für 'Metall-Oxid-Semiconductor', was die Gate-Struktur beschreibt. Inzwischen besteht das Gate nicht mehr aus Metall, sondern aus polykristallinem Silizium, so daßs man allgemeiner auch oft IGFET für 'Isolated Gate FET' sagt.

MOS

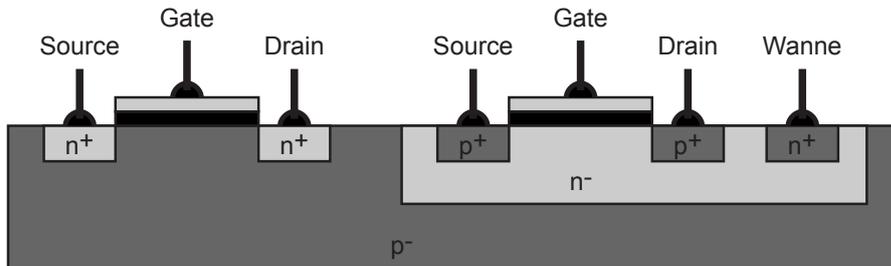


Abbildung 9: Schematischer Aufbau eines n-Kanal MOSFET (links) und eines p-Kanal MOSFET (rechts)

Der Aufbau von MOSFETs ist in Abb. 9 schematisch dargestellt für den Fall, daßs das Grundmaterial, das Substrat schwach p-dotiert ist. Der n-Kanal Transistor ('NMOS') besteht aus n⁺-dotierten Drain- und Source-Gebieten, zwischen denen sich der Kanal ausbilden soll. Die Oberfläche zwischen Source und Drain ist durch ein dünnes Oxid (meist SiO₂) vom Gate isoliert ('IGFET'), das meist aus dotiertem Polysilizium besteht. Das Gatematerial spielt in unserer weiteren Betrachtung keine Rolle (in der Praxis schon!). Der rechts dargestellte PMOS befindet sich in einem schwach n⁻-dotierten Gebiet, der n-Wanne (englisch 'n-well'), die über einen eigenen n⁺-Kontakt angeschlossen ist. Drain und Source sind hier p⁺-Gebiete.

Substrat
NMOS

PMOS
Wanne

3.1 Akkumulation, Verarmung, Inversion

Wir betrachten zunächst eine MOS Struktur (Metall - Isolator - Halbleiter) auf p-dotiertem Silizium, wie das beim NMOS der Fall ist. Je nach der Spannung des Gates unterscheidet man drei Bereiche

1. *Akkumulation*: Bei negativer Gatespannung werden die im Substrat reichlich vorhandenen positiv geladenen Löcher unter das Gate gezogen. Löcher sind frei verfügbar, so daßs sich sehr leicht viele Löcher (je nach Gatespannung) 'akkumulieren' lassen. Unter dem Gate befindet sich also eine leitende Schicht. Betrachtet man die Struktur als Plattenkondensator, so ist die Kapazität pro Einheitsfläche jetzt einfach durch

Akkumulation

$$C_{ox} = \frac{\epsilon_0 \epsilon_{SiO_2}}{t_{ox}} \quad \text{Einheit: } \frac{F}{m^2} \quad (66)$$

gegeben. Da Löcher jederzeit zu- und abgeführt werden können, ist dieser Wert bei hohen und niedrigen Frequenzen gültig.

2. *Verarmung*: Bei zunehmend positiver Gatespannung werden die Löcher von der Grenzschicht weggedrückt (man kann auch sagen Elektronen werden angezogen und besetzen die freien Bindungsstellen der Akzeptor-Atome). Es bildet sich eine *Raumladungszone* (Verarmungszone) aus, deren Dicke mit steigender Spannung ansteigt. Direkt unter dem Gate sind also *keine* beweglichen Ladungsträger mehr vorhanden. Der Plattenabstand steigt und die Kapazität sinkt. Der Wert kann als Serienschaltung von C_{ox} und der Kapazität der Verarmungszone C_{dep} = ε₀ ε_{Si} / t_{dep} berechnet werden.

Verarmung

Raumladungszone

Schwellenspannung

3. *Inversion*: Bei stark positiver Spannung werden weitere Elektronen unter das Gate gezogen. Sie bilden eine dünne, negativ geladene Schicht mit *Minoritätsträgern*. Inversion ist erreicht, wenn die Elektronendichte grösser ist als die Akzeptorendichte. Ab dann sind 'viele' bewegliche Ladungsträger vorhanden. Die zugehörige Gatespannung nennt man die *Schwellenspannung* V_T ('Threshold Voltage'). Die Kapazität steigt wieder auf den Wert des Plattenkondensators mit dünnem Oxid. Bei Spannungsänderungen am Gate müssen Elektronen zu- und abfließen. Da diese aus dem p-Material nicht einfach beschafft werden können, gilt die hohe Kapazität nur bei *niedrigen Frequenzen*. Bei hohen Frequenzen bleibt die Kapazität klein (s. Abb. 10).

Der Transistor wird meist im Bereich *starker Inversion* ('strong inversion') betrieben, also wenn $V_G > V_T$. Ein Betrieb in *schwacher Inversion* ('weak inversion') ist jedoch auch möglich und wird im Low-Power-Design zunehmend eingesetzt. Man spricht dann auch von 'Subthreshold Operation'.

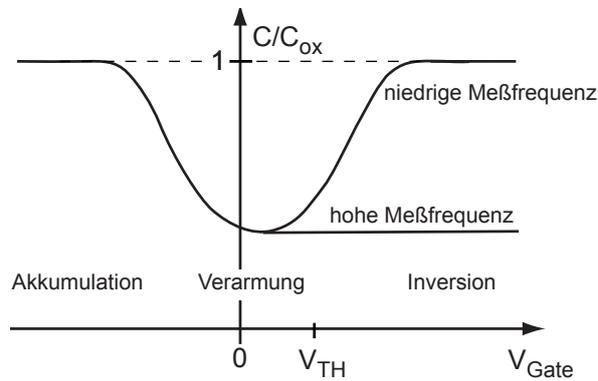


Abbildung 10: Kapazität einer NMOS Struktur als Funktion der angelegten Spannung

3.2 Berechnung der NMOS Kennlinie in starker Inversion

Sobald die Gatespannung über der Schwellenspannung liegt bildet sich eine Schicht aus beweglichen Ladungsträgern unter dem Gate. Die Anzahl Elektronen pro Fläche hängt von der Gatespannung und dem Potential im Kanal ab. Für die Herleitung setzen wir wieder das Sourcepotential auf Null: $V_S = 0\text{V}$. Im Abstand x von der Source (die Geometrie ist wie in Abb. 7) ist die Flächenladung (Einheit: C/m^2) im Kanal

$$Q(x) = C_{ox}(V_G - V_T - V(x)) \quad (67)$$

Hier ist C_{ox} die Flächenkapazität des Gates (66), die nur von der Dicke der Oxidschicht und deren Dielektrizitätskonstante abhängt (Plattenkondensator!). Für $V(x) = 0$ sind also erst ab einer Gatespannung von $V_G = V_T$ negative Ladungsträger im Kanal vorhanden, wie dies ja im vorigen Abschnitt beschrieben wurde. Wenn das Potential im Kanal positiver ist als an der Source, so muß also nach (67) eine höhere Gatespannung angelegt werden um den Kanal auszubilden. Die Ladung im Kanal besteht aus beweglichen Minoritäts-Ladungsträgern (hier Elektronen), die den Kanalwiderstand wesentlich beeinflussen. Ein Stück Kanal der Länge dx hat den Widerstand

$$dR(x) = \frac{dx}{\mu W Q(x)} \quad (68)$$

In diese Gleichung geht auch die Beweglichkeit (Mobilität) μ der Ladungsträger und die Breite W des Kanals ein. Bei einem (von der Koordinate x unabhängigen) Kanalstrom I_D ist der Spannungsabfall über diesem Stück Kanal

$$\begin{aligned} dV(x) &= I_D dR(x) = \frac{I_D dx}{\mu W Q(x)} \\ \text{oder } I_D dx &= \mu W Q(x) dV(x) \end{aligned} \quad (69)$$

Diesen Ausdruck kann man von 0 bis zum Ende des Kanals bei L integrieren:

$$\begin{aligned} \int_0^L I_D dx &= \mu W \int_0^{V_D} Q(x) dV(x) \\ L I_D &= \mu W C_{ox} \int_0^{V_D} (V_G - V_T - V) dV \\ I_D &= \mu C_{ox} \frac{W}{L} \left[(V_G - V_T) V \Big|_0^{V_D} - \frac{1}{2} V^2 \Big|_0^{V_D} \right] \\ I_D &= K \frac{W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \end{aligned} \quad (70)$$

Diese wichtige Gleichung beschreibt den Zusammenhang zwischen Drainstrom und Gate- und Drain-Spannung. In der letzten Zeile wurde der wichtige *Transkonduktanzparameter*

$$K := \mu C_{ox} = \mu \frac{\epsilon_0 \epsilon_{SiO_2}}{t_{ox}} \quad (71)$$

eingeführt, der für n-Kanal Transistoren etwa $K_n = 100 \mu\text{A V}^{-2}$ und, wegen der geringeren Mobilität der Löcher, für p-Kanal Transistoren etwa $K_p = 40 \mu\text{A V}^{-2}$ ist. In Technologien mit kurzer Gatelänge wird meist auch das Gateoxid dünner und K steigt daher.

Für kleine Drainspannungen (genauer für $V_{DS} \ll V_{GS} - V_T$) kann man in (70) den quadratischen Teil vernachlässigen und bekommt so das Verhalten im *linearen Bereich*

$$I_D = K \frac{W}{L} (V_{GS} - V_T) V_{DS} \quad \text{für } V_{DS} \ll V_{GS} - V_T \quad (72)$$

Der Transistor verhält sich also wie ein gesteuerter Widerstand mit

$$R = \frac{V_{DS}}{I_D} = \frac{1}{K \frac{W}{L} (V_{GS} - V_T)} \quad (73)$$

Der Widerstand wird unendlich, wenn V_{GS} die Schwellenspannung erreicht.

linearer Bereich

Widerstand

Gleichung 70 ist nur gültig, solange der Kanal noch nicht abgeschnürt ist. Dies passiert, wenn an der Drain keine Inversionsschicht mehr existiert, also nach (67) wenn

$$V_D = V_{DSat} = V_G - V_T \quad (74)$$

Setzt man das in (70) ein, so erhält man den Drainstrom in Sättigung

$$I_D = \frac{K}{2} \frac{W}{L} (V_{GS} - V_T)^2 \quad (75)$$

Dieser Strom fließt näherungsweise auch für höhere Drainspannungen $V_D > V_{DSat}$, da diese über der Abschnürregion abfallen und die Verhältnisse im Kanal sich nur wenig ändern.

3.3 Kanallängenmodulation

In Realität wächst die Abschnürregion mit steigender Drainspannung jedoch langsam an und der Kanal wird effektiv kürzer. Der Drainstrom steigt dadurch leicht an. Diese *Kanallängenmodulation* führt zu einem endlichen Ausgangswiderstand. Er ist um so höher ('besser'), je länger der Transistor ist (großes L), da sich dann die Vergrößerung der Abschnürregion weniger auswirkt. Man parametrisiert die Kanallängenmodulation in meist guter Näherung durch einen multiplikativen Anteil:

$$I_D = \frac{\mu C_{ox}}{2} \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \quad (76)$$

Die unterschiedlichen Betriebsbereiche eines NMOS sind in Abb. 11 dargestellt. Dort ist auch angedeutet, daß die Verlängerungen aller Sättigungskurven zu negativen Drainspannungen hin die x-Achse bei $-1/\lambda$ schneiden. Man nennt diese Spannung in Anlehnung an den Bipolartransistor auch die 'Early-Spannung'.

Early-Spannung

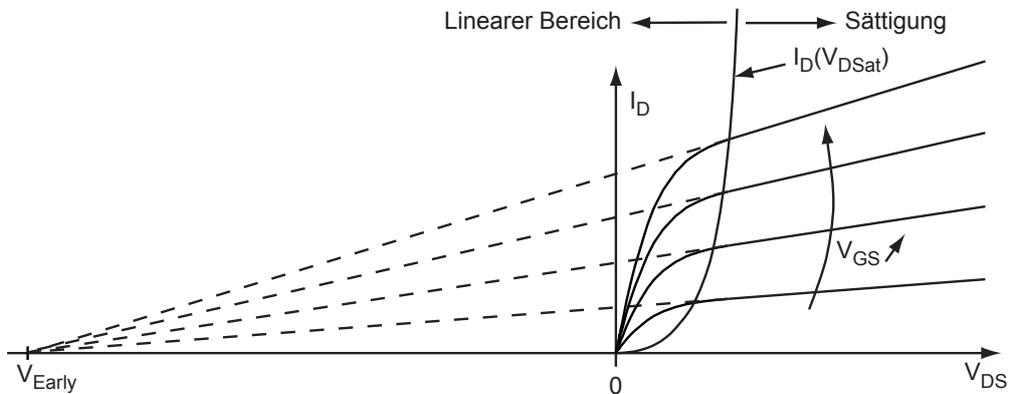


Abbildung 11: Idealisiertes Ausgangs-Kennlinienfeld eines n-Kanal MOSFET

3.4 Substrateffekt

Wir sind bisher davon ausgegangen, daß Source und Substrat des NMOS auf Masse liegen. In vielen Schaltungen ist das Substrat (der 'Bulk') jedoch negativer also die Source (genauer: das Substrat liegt auf Masse und die Source ist positiver.) Die negative Spannung am p-Substrat vergrößert die 'Verarmungszone' unterhalb des Kanals und der Strom wird etwas kleiner. Man kann dies durch einen JFET modellieren, der parallel zur Kanal liegt. Es gibt verschiedene Formeln, die diesen *Substrat-Effekt* (engl. 'body effect') beschreiben. Man kann ihn z.B. durch eine Änderung der Schwellenspannung modellieren:

Substrat-Effekt

$$V_T = V_{T0} + \gamma \left(\sqrt{2\Phi_F - V_{BS}} - \sqrt{2\Phi_F} \right) \quad (77)$$

Hier ist V_{T0} die Schwellenspannung für $V_{BS} = 0\text{V}$, d.h. für $V_S = V_B$. Φ_F ist der Abstand des Ferminiveaus des dotierten Halbleiters vom intrinsischen Fermi-Niveau des undotierten Halbleiters (Φ_F hängt also nur von der Substratdotierung ab). $\gamma = \sqrt{2N_A q \epsilon_{Si} / C_{ox}}$ ist der 'body factor', der die Sensitivität der Schwelle von der Substratspannung beschreibt. Die effektive Schwellenspannung steigt an, wenn das Substrat negativer wird. (Dies ist z.B. der Grund, weswegen die Verstärkung eines Source-Folgers oft wesentlich kleiner als Eins ist.)

3.5 Subthreshold Betrieb

Bisher sind wir davon ausgegangen, daß unterhalb der Schwellenspannung kein Drainstrom fließt. Im 'Subthreshold' Bereich ist die Inversion noch nicht stark ausgeprägt, man spricht daher auch von schwacher Inversion ('weak Inversion') im Gegensatz zur bisher behandelten starken Inversion ('strong inversion'). In diesem Bereich fließt ein kleiner Strom, der exponentiell von der Gatespannung abhängt. Der Ausdruck für den Drainstrom lautet

$$I_D = I_{D0} \frac{W}{L} e^{\frac{V_{GS}}{nV_{th}}} \left(1 - e^{-\frac{V_{DS}}{V_{th}}} \right) e^{\frac{n-1}{n} \frac{V_{BS}}{V_{th}}} \quad (78)$$

Hier ist explizit auch die Abhängigkeit vom Substratpotential beschrieben. I_{D0} ist ein technologieabhängiger Parameter. $V_{th} = kT/q$ ist die Temperaturspannung. Der Parameter n liegt in der Größenordnung von 1.5. Er bestimmt unter anderem wie steil die $\ln[I_D(V_{GS})]$ -Kurve ist und wird daher 'subthreshold slope factor' genannt.

Es ist bemerkenswert, daß in schwacher Inversion die Sättigungsspannung nicht von der Gatespannung abhängt und sehr niedrig ist. Sie beträgt nur wenige V_{th} , also $< 200\text{mV}$.

3.6 Berechnung von C_{GS} in Sättigung

Die Kapazität zwischen Gate und Drain ist in Sättigung klein, da an der Drainseite dann ja kein Kanal mehr vorhanden ist. Um die Kapazität zwischen Gate und Source zu berechnen, müssen wir ermitteln, wie die Gesamtladung im Kanal für eine vorgegebene Gate-Source-Spannung ist. Die Ladung in einem Stück Kanal der Länge dx und der Breite W ist

$$\begin{aligned} dQ(x) &= W Q(x) dx \\ \stackrel{(69)}{=} & \frac{\mu W^2}{I_D} Q^2(x) dV \\ \stackrel{(32)}{=} & \frac{\mu W^2 C_{ox}^2}{I_D} (V_{GS} - V_T - V)^2 dV \\ \stackrel{(75)}{=} & \frac{2 WL C_{ox}}{(V_{GS} - V_T)^2} (V_{GS} - V_T - V)^2 dV \end{aligned} \quad (79)$$

wobei die angegebenen Gleichungen benutzt wurden. Im letzten Schritt wurde der Sättigungsstrom (75) eingesetzt und $K = \mu C_{ox}$ benutzt. Diese Gleichung kann nun z.B. für $V_S = 0$ integriert werden. Das Integral läuft dann von 0 bis zur Spannung $V_G - V_T$ am Ende des Kanals:

$$\begin{aligned} Q_{Kanal} &= \frac{2 WL C_{ox}}{(V_G - V_T)^2} \int_0^{V_G - V_T} (V_G - V_T - V)^2 dV \\ &= \frac{2 WL C_{ox}}{(V_G - V_T)^2} \frac{(V_G - V_T)^3}{3} \\ &= \frac{2}{3} WL C_{ox} (V_G - V_T) \end{aligned} \quad (80)$$

Die Gate-Source Kapazität in Sättigung ist also

$$C_{GS} = \frac{\partial Q_{Kanal}}{\partial V_G} = \frac{2}{3} WL C_{ox} \quad (81)$$