

Metal-Oxide-Semiconductor Transistoren

Stephan Epp
hjstephan86@gmail.com

16. Juli 2025

Inhaltsverzeichnis

1	Einleitung	2
2	MOS Transistor	2
2.1	Sperrbereich	2
2.2	Triodenbereich / Linearer Bereich	2
2.3	Sättigungsbereich	3
2.4	U-I-Kennlinie für $U_{GS} = 2\text{ V}$	3
3	Complementary Metal-Oxide-Semiconductor (CMOS)	4
4	Logische Gatter: NND und NOR	5
5	Eingebettetes System	6
5.1	Scheduling	6

1 Einleitung

Metal-Oxide-Semiconductor (MOS) Transistoren bilden die Grundlage für hochintegrierte Schaltungen. Die U-I-Kennlinie eines nMOS-Transistors folgt in den verschiedenen Betriebsbereichen spezifischen mathematischen Formeln. Diese Formeln basieren auf physikalischen Modellen des Transistors und beschreiben den Drain-Source-Strom I_{DS} in Abhängigkeit von den angelegten Spannungen. Diese bestehen hauptsächlich aus der Gate-Source-Spannung U_{GS} und der Drain-Source-Spannung U_{DS} . Es gibt drei Hauptbetriebsbereiche für einen nMOS-Transistor: den Sperrbereich, den Triodenbereich und den Sättigungsbereich. Das n vor MOS bezieht sich dabei auf das Substrat des nMOS Transistors. Der nMOS Transistor besteht aus einem n-p-n-Übergang, bei dem ein leitender Zustand nur dann erfolgt, wenn Elektronen als Minoritätsladungsträger im p-Substrat durch den Feldeffekt unter die Kondensatorplatte gezogen werden.

2 MOS Transistor

2.1 Sperrbereich

In diesem Bereich ist der Transistor ausgeschaltet und leitet praktisch keinen Strom.

- **Bedingung:** $U_{GS} < U_{TH}$ (wobei U_{TH} die Schwellenspannung ist).
- **Formel:**

$$I_{DS} \approx 0$$

- **Erklärung:** Wenn die Gate-Source-Spannung U_{GS} unter der Schwellenspannung U_{TH} liegt, kann sich kein leitender Kanal zwischen Source und Drain bilden, und es fließt kein nennenswerter Strom, d.h., $I_{DS} \approx 0$. Idealisiert ist $I_{DS} = 0$, in der Realität gibt es einen sehr kleinen Leckstrom (Subthreshold-Strom), der oft vernachlässigt wird.

2.2 Triodenbereich / Linearer Bereich

Dieser Bereich wird auch als ohmscher Bereich bezeichnet, da der Transistor hier wie ein spannungsgesteuerter Widerstand wirkt. Der Strom steigt hier nahezu linear mit U_{DS} an (für kleine U_{DS}) und ist stark von U_{GS} abhängig.

- **Bedingung:** $U_{GS} > U_{TH}$ und $U_{DS} < (U_{GS} - U_{TH})$
- **Formel (vereinfachtes Modell für lange Kanäle):**

$$I_{DS} = \mu_n C_{ox} \frac{W}{L} \left((U_{GS} - U_{TH}) U_{DS} - \frac{1}{2} U_{DS}^2 \right)$$

- μ_n : Elektronenbeweglichkeit im Kanal
- C_{ox} : Oxidkapazität pro Flächeneinheit
- W : Kanalbreite
- L : Kanallänge
- U_{TH} : Schwellenspannung
- **Erklärung:** Wenn U_{GS} über U_{TH} liegt, bildet sich ein Kanal. Solange U_{DS} nicht zu hoch ist, verhält sich der Kanal wie ein variabler Widerstand. Der Strom nimmt mit zunehmendem U_{DS} zu und die Breite des Kanals verringert sich zum Drain hin leicht, was zu einer nicht-linearen Abhängigkeit führt, die durch den Term U_{DS}^2 beschrieben wird. Der nicht-lineare Anstieg vom gesperrten Zustand im Übergang bis in den leitenden Zustand bezieht sich auf den Übergang vom Sperrbereich in diesen Bereich und den anfänglich nicht-linearen Anstieg des Stroms, bevor er in die Sättigung geht.

2.3 Sättigungsbereich

In diesem Bereich verhält sich der Transistor wie eine spannungsgesteuerte Stromquelle, da der Drain-Source-Strom weitgehend unabhängig von U_{DS} ist.

- **Bedingung:** $U_{GS} > U_{TH}$ und $U_{DS} \geq (U_{GS} - U_{TH})$

- **Formel (vereinfachtes Modell für lange Kanäle):**

$$I_{DS} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (U_{GS} - U_{TH})^2 (1 + \lambda U_{DS})$$

- Der Term $(1 + \lambda U_{DS})$ berücksichtigt den **Kanallängenmodulationseffekt**, λ ist der Kanallängenmodulationsparameter, der eine leichte Zunahme des Stroms mit U_{DS} in der Sättigung beschreibt. Ohne diesen Effekt wäre der Strom konstant.

- **Erklärung:** Sobald U_{DS} einen bestimmten Wert erreicht ($U_{DS,sat} = U_{GS} - U_{TH}$), wird der Kanal am Drain-Ende abgeschnürt (pinch-off). Eine weitere Erhöhung von U_{DS} führt nicht zu einer signifikanten Zunahme des Stroms, da der Stromfluss durch die Anzahl der Ladungsträger im Kanal und die Gate-Spannung begrenzt wird.

2.4 U-I-Kennlinie für $U_{GS} = 2\text{ V}$

Für $U_{GS} = 2\text{ V}$, $U_{TH} = 0.2\text{ V}$ und $k = 1\text{ mA/V}^2$ ergibt sich:

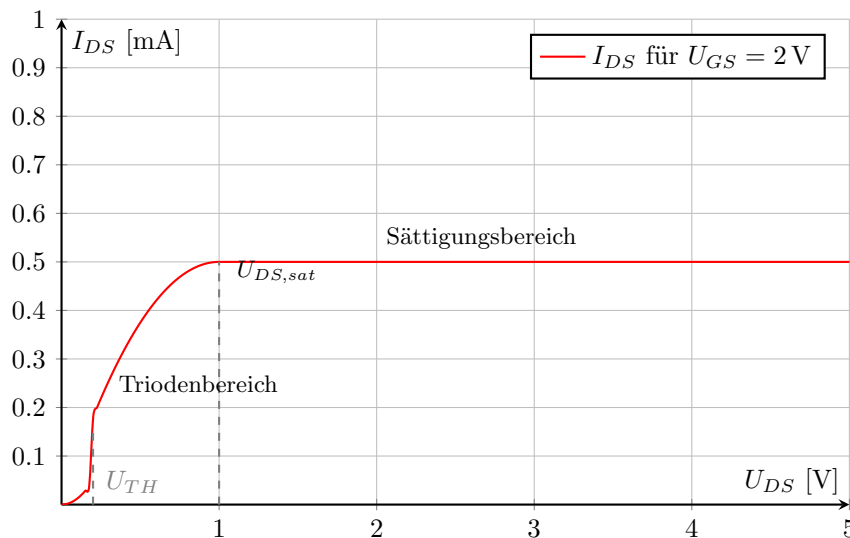


Abbildung 1: Idealisierte U-I-Kennlinie eines nMOS-Transistors mit $U_{GS} = 2\text{ V}$

- **Sperrbereich** ($U_{GS} < U_{TH}$, näherungsweise $U_{DS} < 0.2\text{ V}$):

$$I_{DS} \approx a \cdot U_{DS}^2 \quad \text{mit } a = 1.25\text{ mA/V}^2$$

- **Triodenbereich** ($0.2\text{ V} < U_{DS} < 1\text{ V}$):

$$I_{DS} = k \left((U_{GS} - U_{TH}) U_{DS} - \frac{1}{2} U_{DS}^2 \right)$$

- **Sättigungsbereich** ($U_{DS} \geq 1\text{ V}$):

$$I_{DS} = \frac{1}{2} k (U_{GS} - U_{TH})^2 = 0.5\text{ mA}$$

3 Complementary Metal-Oxide-Semiconductor (CMOS)

Bevor der CMOS Schaltkreis betrachtet wird, wird das Schaltsymbol des pMOS-Transistors und das Schaltsymbol des nMOS-Transistors in Abbildung 2 gezeigt.



Abbildung 2: Schaltsymbol des pMOS-Transistors P_1 und des nMOS-Transistors N_1

Complementary Metal-Oxide-Semiconductor (CMOS) Schaltkreise sind Schaltkreise, in denen der pMOS- und der nMOS-Transistor komplementär zueinander verbunden werden. Abbildung 3 zeigt eine CMOS-Schaltung mit der Betriebsspannung U , der Eingangsspan-

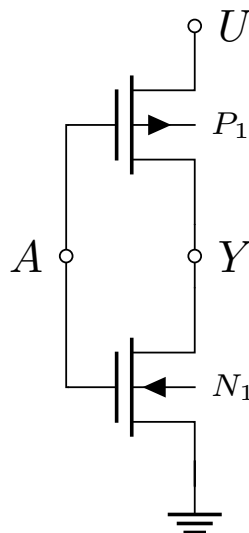


Abbildung 3: CMOS-Schaltung

nung A für die Transistoren P_1 und N_1 und der Ausgangsspannung Y . Dabei ist P_1 der pMOS-Transistor und N_1 der nMOS-Transistor.

Beim nMOS-Transistor N_1 zeigt der Pfeil auf den Transistor. Dies zeigt seine physikalische Wirkungsweise. Ist die Eingangsspannung A ausreichend groß, dann liegt am Gate eine positive Spannung an. Das bewirkt wie beim Kondensator einen Feldeffekt zwischen den beiden Kondensatorplatten, der die Minoritätsladungsträger (Elektronen) im p-Substrat des nMOS-Transistors unter die Oberfläche der gegenüberliegenden Kondensatorplatte zieht. Erst durch diesen Feldeffekt tragen die Minoritätsladungsträger in ihrer Unterzahl im p-Substrat zum Stromfluss bei. Es entsteht ein leitender Kanal von Elektronen. Ist die Eingangsspannung nämlich nicht ausreichend hoch, sperrt der nMOS-Transistor. Daher wird der nMOS-Transistor auch nMOS-Feldeffekttransistor (nMOSFET) genannt.

Beim pMOS-Transistor P_1 zeigt der Pfeil weg von der unteren Kondensatorplatte. Das liegt daran, dass die Elektronen im pMOS-Transistor nicht Minoritätsladungsträger sondern

Majoritätsladungsträger im n-Substrat sind. Das heißt, der pMOS-Transistor leitet sogar schon bei einer Eingangsspannung von $A = 0V$. Der Feldeffekt beim pMOS-Transistor führt dazu, dass bei ausreichender Eingangsspannung A am Gate die Minoritätsladungsträger im n-Substrat unter die Kondensatorplatte gezogen werden und der pMOS-Transistor sperrt. Daher wird der pMOS-Transistor auch pMOS-Feldeffekttransistor (pMOSFET) genannt.

Beide Transistoren P_1 und N_1 in Reihe geschaltet haben somit ein physikalisch komplementäres Sperr- und Leitverhalten zueinander. Dieses komplementäre Sperr- und Leitverhalten gibt diesem Schaltkreis den Namen Complementary Metal-Oxide-Semiconductor (CMOS) Schaltkreis.

Betrachtet man das logische Schaltverhalten der CMOS Schaltung in Abhängigkeit der Betriebsspannung U , der Eingangsspannung A und der Ausgangsspannung Y , wird klar, dass mit dieser Schaltung ein Inverter betrieben wird. Ist $A = 0$, leitet P_1 und N_1 sperrt. Das heißt, das Bezugspotenzial für Y ist U und damit ist $Y = 1$ (Pull-Up-Bezugspotenzial). Ist $A = 1$, sperrt P_1 und N_1 leitet. Das heißt, das Bezugspotenzial für Y ist die Masse und damit ist $Y = 0$ (Pull-Down-Bezugspotenzial).

4 Logische Gatter: NND und NOR

Ein logisches Gatter (oft auch einfach Gatter oder englisch (logic) gate genannt) ist ein grundlegender Baustein in der digitalen Elektronik, der eine bestimmte boolesche Funktion berechnet. In Kapitel 3 wird ein Inverter beschrieben, der die boolesche Funktion

A	B	$A \text{ NND } B$
0	0	1
0	1	1
1	0	1
1	1	0

A	B	$A \text{ NOR } B$
0	0	1
0	1	0
1	0	0
1	1	0

Abbildung 4: Boolesche Funktion NND und NOR

des NOT Gatters in Abhängigkeit des Eingangssignals A berechnet. Andere Gatter, die durch CMOS-Schaltungen gebildet werden können, sind das NAND (kurz: NND) Gatter und das NOR Gatter. Das NND Gatter und das NOR Gatter bilden boolesche Funktionen in Abhängigkeit der Eingangssignale A und B . Abbildung 4 zeigt die vollständig beschriebene boolesche Funktion beider Gatter. Dadurch, dass diese booleschen Funktionen durch MOSFETs gebildet werden können, sind ihre Funktionen als Hardwarelösung nutzbar. Das bedeutet, der Prozessor kann diese Funktion ohne ein geschriebenes Programm direkt in der Hardware auswerten.

Wie lange braucht der Prozessor dafür? Der Takt eines Prozessors ist die kleinste Zeiteinheit, in der der Prozessor periodisch einen Arbeitsschritt ausführt. Die kleinste Zeiteinheit wird aber deutlicher beschrieben durch den Begriff *Slot*, da Slot die dem Prozessor zur Verfügung stehende Zeit betont. Der Vorteil von in Hardware nutzbaren Funktionen ist der, dass der Prozessor in einem Slot z.B. die Ergebnisse der NND Funktion oder der NOR Funktion auswerten kann. Würde der Prozessor z.B. für die NND Funktion ein geschriebenes Programm ausführen, könnte er mit dieser Programmausführung die NND Funktion nicht in einem Slot auswerten.

Es ist zu beachten, dass die NND Funktion oder die NOR Funktion mit der einfachen Realisierung durch MOSFETs jeweils in einem Slot vom Prozessor ausgewertet werden kann. Diese Funktionen gehören zu den grundlegenden Hardwarefunktionen eines Prozessors. Es gibt komplexere Funktionen, die auch durch elektronische Schaltungen in Hardware realisiert werden. Diese Funktionen aber kann der Prozessor nicht in einem Slot auswerten, benötigt dafür aber immer noch weit weniger Slots als wenn er sie durch ein geschriebenes Programm auswerten müsste. Andererseits ist es nicht so, dass deshalb jede komplexere

Funktion in Hardware gelöst wird. Der Grund dafür ist, dass der Aufwand bzw. die Kosten und der Nutzen unverhältnismäßig sind im Vergleich dazu, dass die komplexeren Funktionen in einem geschriebenen Programmen ausgewertet werden. Die komplexen Funktionen sind heute so groß und ändern sich so häufig, dass es generische Microcontroller braucht, die programmierbar sind mit geschriebenen Programmen. Diese komplexen Funktionen werden von eingebetteten Systemen ausgeführt wie z.B. die Auswertung von Radarobjekten in der Automobilindustrie.

5 Eingebettetes System

Ein eingebettetes System ist eine Hardware-/Softwareeinheit, die innerhalb einer physikalischen Umgebung Aufgaben in Echtzeit bearbeitet. Die Aufgaben oder Tasks haben harte Deadlines und müssen vom Prozessor so abgearbeitet werden, dass alle Deadlines eingehalten werden. Eine harte Deadline darf auf keinen Fall überschritten werden. Beispiele für physikalische Umgebungen, in denen eingebettete Systeme heute eingebettet sind, sind zum Beispiel Flugzeuge, Züge, Autos oder Smartphones.

5.1 Scheduling

Damit alle Tasks in einer geplanten und geordneten Reihenfolge abgearbeitet werden können, gibt es den Scheduler, der den Schedule für alle Tasks plant. An diesen Schedule hält sich der Prozessor zur Bearbeitung aller Aufgaben, da erst so das Einhalten aller Deadlines *garantiert* wird.

Earliest Deadline First (EDF) ist eine Vorgehensweise zur Scheduleberechnung, bei der die Menge an Tasks vorher bekannt ist. Nach EDF wird die Menge an Tasks gierig nach Deadlines sortiert und entsprechend priorisiert. Je früher die Deadline einer Task ist, desto höher ist ihre Priorität und desto früher wird sie ausgeführt. Es kann passieren, dass während der Ausführung des Schedules eine Task blockiert, da sie auf eine Ressource warten muss. Die Idee von EDF⁺ ist, dass Tasks, die blockieren, mit geringster Priorität bestraft werden und anschließend der EDF Schedule so schnell wie möglich neu berechnet und dann ausgeführt wird.

Satz 5.1. *EDF⁺ ist optimal in der Garantienaussage über die gesamte Zeit, in der Tasks ausgeführt werden.*

Beweis. EDF ist optimal in der Auslastung des Prozessors. Dies ist eine Garantienaussage. Sobald eine Task verdrängt werden muss, weil sie blockiert und auf eine andere Ressource wartet, und damit vom optimalen Schedule abgewichen werden muss, ist die Garantienaussage verletzt. Das heißt, es muss dann so schnell wie möglich die Garantienaussage wiederhergestellt werden. Dazu wird die Task, die blockiert hat, mit niedrigster Priorität bestraft. Anschließend wird EDF zur Scheduleberechnung ausgeführt und dieser neue Schedule mit der Garantienaussage für alle Tasks zur Bearbeitung genutzt. Dieses Vorgehen ist optimal hinsichtlich der Zeit der Ausführung des Schedules, über die eine Garantienaussage des optimalen EDF Schedules gemacht wird. Denn das Risiko, dass die blockierte Task noch einmal blockiert, wird dadurch minimiert, dass sie, weil sie blockiert hat, die niedrigste Priorität erhält. □

Es ist zu beachten, dass es während der Ausführung von Tasks, die nach dem EDF⁺ Schedule abgearbeitet werden, immer wieder zu Unterbrechungen kommen kann, wenn das Nutzen anderer Ressourcen nicht möglich ist. Das heißt, die Anzahl der Unterbrechungen sind vorher nicht bekannt und können dazu führen, dass einige Tasks aus der Menge aller Tasks *nicht* bis zu ihrer Deadline abgearbeitet werden können.