

Vorlesung 3

Folie 1

Thema dieser Vorlesung sind die CMOS Gatter

In einer CMOS Technologie gibt es zwei Hauptbauteile NMOS und PMOS

CMOS bedeutet Komplementäre Metal-Oxid-Semiconductor Transistoren.

Der Name ist veraltet da heute die Gate-Elektrode nicht aus Metall ist.

Folie 3

Die Transistorarten haben die folgenden Kontakte:

Source, Drain. Sie werden durch Diffusion erzeugt

Source und Drain sind in digitalen Schaltungen möglichst nah einander. Sie haben eine andere Dotierung als der Bereich zwischen ihnen.

Die Dotierung von Source und Drain bestimmt die Ladungsträger – NMOS benutzt n-dotierte Source und Drain – Ladungsträger sind Elektronen.

PMOS benutzt p-dotierte Source und Drain – Ladungsträger sind Löcher.

Folie 4

Ein NMOS leitet Strom (Elektronen) wenn man durch eine positive Gate-Source Spannung im P-Bereich zwischen dem Source und dem Drain eine Typinversion erreicht. Das heißt, die Löcher im P-Bereich werden durch das E-Feld verdrängt und ein Elektronkanal an der Silizium/Oxid Oberfläche erzeugt.

Hier ist das folgende wichtig: NMOS leitet auch wenn der Drain ein höheres Potential als das Gate hat. Wichtig ist nur das Gate Potential in Bezug auf die Source höher ist.

Hier definiert man die Transistorschwelle als die minimale Gate-Source Spannung die den Transistor einschaltet. Diese Spannung ist fast immer im Bereich 0.3 – 0.5V.

Folie 3 zeigt die NMOS Struktur. Wir sehen, dass der Transistor symmetrisch ist (im Bezug auf Source und Drain).

Wie wissen wir dann, welcher Kontakt Drain und welcher Source ist?

Eigentlich können Source und Drain die Rollen auch vertauschen.

Es gilt: für ein NMOS ist die Source der Kontakt mit niedrigerem Potential.

Folie 6 (links)

Betrachten wir einige Beispiele:

Annahmen:

Wir haben auf einem Chip die Masse-Leitung (GND), deren Potential null ist und eine Versorgungsspannung (VDD) mit dem Potential 1.2V.

Die NMOS - Transistoreschwelle ist 0.4V

Wenn Source des NMOS – Transistors an GND angeschlossen ist, leitet er wenn seine Gate Elektrode an VDD-Potential liegt. Gate Source Spannung ist 1.2V also höher als die Schwelle.

Folie 6 (rechts)

Hier haben wir zwei NMOS Transistoren in Serie, und die Source vom T1 ist GND.

Wir haben den Drain des Transistors T2 an einen Pull-Up Widerstand (R) angeschlossen – dieser Widerstand verbindet die Transistoren und VDD.

Beide Transistor-Gates sind VDD.

Der untere Transistor (T1) leitet da $V_{gs1} > \text{Schwelle}$. Der obere Transistor (T2) leitet nur wenn seine $V_{gs2} > 0.4 \text{ V}$ ist. Dies ist gewährleistet, falls das Potential

V_x klein genug ist, also falls der Widerstand des leitenden Transistors deutlich kleiner als R-Pullup ist ($R(T1) \ll R$). Denn, es fließt ein Strom durch die zwei Transistoren und den Widerstand R, und es bildet sich eine Spannung zwischen den Drain und Source vom T1. Für $R(T1) \ll R$ ist diese Spannung im Bereich $\sim 100\text{mV}$ und die V_{gs} des oberen Transistors ist größer als die Schwelle V_{th} .

Folie 7

Betrachten wir jetzt den Fall, wenn das Gate Potential des unteren Transistors T1 null ist. Das Gate-Potential des oberen Transistors ist VDD. Die Frage ist – welcher Transistor leitet und welcher nicht?

T1 leitet nicht da seine $V_{gs} = 0$. Da der Transistor nicht leitet, kann kein Strom fließen -> das Potential am Drain vom T2 ist VDD.

Transistor T2 ist auf dem Umschaltunkt zwischen dem leitenden- und gesperrten Zustand. Wenn der Transistor T2 leiten würde, wäre sein Source Potential VDD. In dem Fall wäre V_{gs2} null – also die Annahme T2 leitet ist falsch. Andersfalls, wenn der Transistor nicht leitet ist das Source Potential undefiniert.

Wir können daraus schließen, dass sich NMOS Transistoren besser (bzw. definiert) verhalten, wenn ihre Source-Kontakte an GND angeschlossen werden (wie bei T1).

Folien 8-11

Wir zeigen zwei (schlechte) Beispiele wo NMOS falsch verwendet wurde:

Wir könnten versuchen die NMOS-Transistoren als Schalter mit einem Pull-Down Widerstand zu kombinieren.

Die Einfachste Möglichkeit wäre ein ODER-Gate – Folie 8.

Wenn beide Gate Potentiale null sind, leiten die Transistoren nicht und der Ausgang ist null. Das ist in Ordnung.

Wenn mindestens ein Gate VDD wäre (T1) (Folie 9), würden wir annehmen, dass der Transistor T1 leitet und, dass der Ausgang VDD ist. Das ist nicht ganz korrekt. Denn mit VDD am Ausgang, wäre auch die VGS des „leitenden“ Transistors T1 null. In Wirklichkeit, also, haben wir am Ausgang eine Spannung, die kleiner als $V_{DD} - V_{th}$ ist. Das wäre unter Umständen nicht für eine logische Eins ausreichend.

Darüber hinaus, wäre es mit NMOS Transistoren und Pull-Down Widerständen unmöglich Inverter zu realisieren. Wir haben gesehen, dass die Inverter für die die logischen Funktionen wichtig sind.

Folien 10 - 11

Es gibt eine wichtige logische Schaltung, die sich mit idealen Schaltern einfach realisieren lässt – der Multiplexer.

Folie 10 zeigt den Aufbau des Multiplexers.

Auch hier wäre ein NMOS Schalter nicht ausreichend. Wenn wir am Eingang logische Eins = VDD haben (Folie 11), und wenn der NMOS Gate VDD ist, wäre der Ausgang in Realität $\sim V_{DD} - V_{th}$ oder niedriger.

Alles was für ein NMOS in den letzten Folien gesagt wurde, gilt auch für einen PMOS, mit dem Unterschied, dass ein PMOS besser funktioniert, wenn sein Source an VDD angeschlossen ist, und dass ein PMOS leitet, wenn seine Gate - Source Spannung negativer als eine (negative) Schwelle ist. PMOS Transistoren kann man mit deshalb einem Pull-Down Widerstand kombinieren.

Folien 12 und 13

Folie 13 zeigt den Aufbau eines PMOS Transistors (neben einem NMOS).

PMOS hat p-dotierte Source und Drain Bereiche. Der Bereich zwischen den Source und Drain (Bulk/Kanalbereich) ist n-Dotiert. Eine negative Gate - Source Spannung erzeugt die Typinversion im Kanalbereich. Die Elektronen werden abgestoßen und die Löcher an die Silizium/Oxid Oberfläche angezogen. So kann

der Transistor leiten, die Löcher können den Bereich zwischen Drain und Source überqueren.

Im PMOS ist Source der Kontakt mit höherem Potential.

Folie 14

Wie gesagt, ein PMOS leitet besser wenn sein Source an VDD angeschlossen ist.

PMOS Transistoren kann man mit Pull-Down Widerständen kombinieren.

Auf diese Weise kann ein PMOS Inverter realisiert werden.

Folie 15

Wenn das Gate-Potential VDD ist, ist $V_{gs} = 0$ -> der Transistor leitet nicht und die Ausgangsspannung ist null.

Folie 16

Wenn das Gate Potential GND ist, ist $V_{gs} = -V_{DD}$, was kleiner als die Schwelle ist. (Die Schwelle ist typischerweise $-0.4V$.) Der Transistor leitet und die Ausgangsspannung ist $\sim V_{DD}$. Hier nehmen ich an, dass der Widerstand des leitenden Transistors deutlich niedriger als R ist.

Folie 17

Logische Schaltungen mit NMOS-Transistoren und Pullup-Widerständen und mit PMOS-Transistoren und Pulldown Widerständen sind möglich.

Auf diesem Prinzip arbeitet die RTL-Logik „Familie“.

Diese Schaltungen haben drei Nachteile:

Bei den eingeschalteten Transistoren fließt Strom, wir haben einen „DC-Stromverbrauch“. (Die logischen Gatter verbrauchen Strom auch wenn sie nicht

getaktet werden.) Wir können den Stromverbrauch durch einen größeren Widerstand minimieren, aber dann bekommt man zwei andere Nachteile:

1. Große Polysilizium-Widerstände nehmen auf einem Chip große Fläche.
2. Logische Schaltungen, die auf großen Widerständen basieren sind aus folgendem Grund langsam: Oft werden die logischen Schaltungen an viele andere Gatter angeschlossen. Das führt zu einer kapazitiven Last. Die Geschwindigkeit, mit der diese Last aufgeladen wird, (z.B. wenn sich das logische Niveau von null (GND) auf eins (VDD) ändert) hängt vom $R \cdot C$ Produkt ab. Beim großen R und C, sind die Schaltungen langsam.

Folie 18

Wir haben gesehen, dass NMOS- und PMOS Transistoren im folgendem Sinne komplementär sind: NMOS-Transistoren können eine Leitung ans Potential der Masse gut „ziehen“ und PMOS-Transistoren ans VDD-Potential.

Deswegen lassen sich die Transistoren in logischen Gattern gut kombinieren. Hier hilft auch die Tatsache, dass die Gates verschiedene Polaritäten haben sollen um die Transistoren einzuschalten.

Die Idee ist z.B. bei einem NMOS RTL Inverter, den Pullup-Widerstand durch den PMOS zu ersetzen.

Folie 18 zeigt den CMOS Inverter.

Folien 19 und 20

Die Schaltung funktioniert wie folgend: Wenn der Eingang VDD ist (logische Eins) (Folie 19) leitet der NMOS ($V_{gs} = VDD$) und der PMOS sperrt ($V_{gs} = 0$). Der Ausgang ist GND (logische Null).

Ähnlich, wenn der Eingang GND ist (Folie 20) leitet PMOS und der NMOS sperrt. Der Ausgang ist VDD (logische Eins).

Folie 21

CMOS Gatter haben einige Vorteile gegenüber RTL-Logik.

1. Wenn der Eingang konstante Spannung hat, fließt kein DC-Strom durch den Inverter.
2. Der Inverter besteht nur aus Transistoren, die sich auf dem Chip wenig Fläche nehmen.
3. Die Umlade-Zeit hängt vom Widerstand des leitenden Transistors und der Ausgangskapazität. Die Transistoren können entsprechend dimensioniert werden so, dass die Schaltung schnell ist.

Folie 22

Die Analyse und das Design des Inverters gehören ins Gebiet der analogen Elektronik. Wir werden deshalb einige Methoden der analogen Elektronik in weiterer Analyse verwenden.

Wir werden zuerst die Großsignal-Analyse für langsame Signale beschreiben.

Dann werden wir die AC-Analyse durchführen, um die Geschwindigkeit der Schaltung abzuschätzen.

Die, in der analogen Elektronik übliche, Approximation - Kleinsignalanalyse hilft hier nicht. In den digitalen Schaltungen sind die Signale „groß“ und die Nichtlinearitäten sind unvermeidbar, sogar erwünscht. Bei der AC-Analyse werden wir also die nichtlinearen Kennlinien verwenden.

DC - Analyse.

Ziel: Wir leiten die Eingang-Ausgangs-Kennlinie des Inverters her.

Theorie:

Transistoren werden mit zwei Arten von Kennlinien beschrieben - der Eingangskennlinie ($I_d = f(V_{gs})_{\text{Sättigung}}$) und der Ausgangskennlinie ($I_d = f(V_{ds})$).

Für uns ist besonders die Ausgangskennlinie interessant:

Folie 22 zeigt die Ausgangskennlinie eines NMOS Transistors für drei verschiedene Gate-Source Spannungen. Die Ausgangskennlinie beschreibt den Drain - Source Strom als Funktion der Drain Source Spannung.

Die erste Kennlinie zeigt den Strom für $V_{GS} = 0$. Der Strom ist null für alle V_{ds} Spannung.

Zweite Kennlinie zeigt I_{ds} für $V_{GS} < V_{DD}$.

Die dritte Kennlinie zeigt den Strom für $V_{GS} = V_{DD}$. Die Kennlinie ist eine Parabel, der Strom steigt nicht mehr für $V_{ds} > V_{DD} - V_{th}$.

Folien 23-31 (Theorie)

Wir sehen also, dass sich der Transistor wie ein spannungsabhängiger Widerstand verhält. Sein Widerstand wird größer für größere V_{ds} .

Warum ist das so?

Es wurde erwähnt, dass eine positive Gate - Source Spannung die Elektronen anzieht und einen Kanal zwischen Source und Drain bildet. Dieser Kanal leitet den Strom.

Wenn der Strom im Kanal fließt, gibt es auch eine Spannungsänderung im Kanal. selbst.

Am Kanal-Anfang nahe Source ist das Kanalpotential gleich dem Source Potential.

Am Kanalende nahe Drain ist das Kanalpotential höher. (In den Folien wir ein höheres Potential als höhere Lage dargestellt.) Eine kleinere Gate-Kanal

Spannung nahe Drain führt dazu, dass die Elektronen im Kanal dort weniger die „Anziehungskraft“ der Gate spüren. Deshalb ist die Elektronen-Dichte im Kanalende niedriger und der Widerstand des Kanals höher.

Deshalb auch wird der effektive Widerstand des Kanals höher, wenn mehr Strom durch ihn fließt (bzw. wenn das Drain-Potential erhöht wird). Für einen bestimmten Stromwert, wird der Widerstand am Kanalende unendlich und der Strom kann nicht weiter steigen. Der Kanal wird „abgeschnürt“, der Transistor befindet sich in Sättigung. Der Drain Strom hängt nicht mehr von V_{ds} ab. Dieser Effekt wird in den analogen Schaltungen benutzt, um Verstärker zu realisieren.

In digitaler Elektronik ist dieser Effekt (Sättigung) im Prinzip für die Funktionalität unwichtig, er beeinflusst nur die Umladezeit.

Es kann hergeleitet werden, dass der Transistorstrom mit folgender Gleichung beschrieben werden kann:

$$I_{ds} = \mu C_{ox} W/L ((V_{gs} - V_{th}) V_{ds} - V_{ds}^2 / 2)$$

Die Steigung am Anfang ist $\mu C_{ox} W/L (V_{gs} - V_{th})$ (Folie 24)

Für $V_{gs} < V_{th}$ der Strom ist null – der Transistor sperrt. (Folie 25)

Für $V_{gs} > V_{ds} - V_{th}$ ist der Transistor in Sättigung. (Folie 26) Der Sättigungsstrom ist:

$$I_{ds} = \frac{1}{2} \mu C_{ox} W/L (V_{gs} - V_{th})^2$$

Die Formen kann man wie folgend erklären:

Folie 27

Für höhere V_{gs} Spannung gibt es mehr Elektronen im Kanal – der spezifische Leitwert ist höher, Widerstand kleiner – der Strom steigt.

Deshalb haben wir den Faktor $V_{gs} - V_{th}$.

Folien 28 und 29

Die Faktoren W/L und μ haben wir weil der Widerstand bzw. der Leitwert auch von Mobilität der Elektronen und der Geometrie der Gate-Elektrode abhängig ist. Größere Stromfläche $W * t$ (t ist Dicke des Kanals) bedeutet mehr Strom und ein längerer Kanal (L) weniger Strom.

Folie 30

Faktor $-V_{ds}^2/2$ modelliert die Erhöhung des Widerstands beim V_{ds} Anstieg.

Der Strom steigt immer langsamer.

Folie 31

Faktor C_{ox} ist die Oxidkapazität pro Fläche. So größer diese Kapazität ist, desto mehr Ladung haben wir im Kanal für eine gegebene V_{gs} Spannung.

Die Dicke des Kanals t ist im Faktor $C_{ox} * (V_{gs} - V_{th})$ enthalten, da dieser Faktor die Ladungsdichte $* t$ ergibt. Deshalb taucht die Dicke t in der endgültigen Formel nicht.

Folie 32

Folgendes ist wichtig:

Man kann den Leitwert des Transistors erhöhen indem man W/L erhöht. Man kann den Transistor nach dem Bedarf breiter oder kürzer machen.

Die Mobilität der Löcher ist etwa 2x niedriger, deshalb leitet ein PMOS mit gleichem W/L Verhältnis etwa 2x schlechter. Das soll beim Design berücksichtigt werden.

Folie 33

Die DC-Kennlinie des Inverters bekommt man, wenn man auf einem Graph die Kennlinien $I(V_{ds})$ von PMOS und NMOS zeichnet.

Für den gemeinsamen Graph gilt:

V_{ds} des NMOS Transistors ist gleichzeitig V_{out} .

V_{gs} des NMOS Transistors ist gleichzeitig V_{in} .

V_{ds} (Betrag) des PMOS Transistors ist gleichzeitig $V_{DD} - V_{out}$.

V_{gs} (Betrag) des PMOS ist $V_{DD} - V_{in}$

Schwellen (Betrag) von NMOS und PMOS sind normalerweise gleich.

Wenn wir annehmen, dass W/L des PMOS Transistors etwa 2x größer ist als W/L vom NMOS, sind die Faktoren $\mu C_{ox} W/L$ gleich und die Kennlinien gegen $x=V_{DD}/2$ spiegelsymmetrisch.

Wenn wir annehmen, dass $I_{ds} \text{ NMOS} = I_{ds} \text{ PMOS}$ (es fließt kein DC-Strom nach außen – die Last ist kapazitiv), befindet sich V_{out} im Schnittpunkt beider Kennlinien.

Man kann wie folgend die DC Kennlinie herleiten:

Folien 34-35

Für $V_{in} < V_{th}$, leitet der NMOS nicht und $V_{out} = V_{DD}$

Folien 36-37

Für $V_{in} > V_{DD} - V_{th}$ leitet der PMOS nicht und $V_{out} = 0$.

Folie 38

Im Bereich $V_{th} < V_{in} < V_{DD} - V_{th}$ leiten beide Transistoren

Folie 39

Die Kennlinie ist besonders steil im Bereich wo beide Transistoren in Sättigung sind

Für NMOS ist Sättigung für $V_{ds} > V_{gs} - V_{th}$ gegeben also $V_{out} > V_{in} - V_{th}$

Folie 40

Für PMOS haben wir Sättigung wenn $|V_{ds}| > |V_{gs}| - |V_{th}|$ also $V_{out} < V_{in} + V_{th}$

Sättigungsbereich ist im Graph – Folien 39-40 gekennzeichnet.

Folien 41-42

Die Kennlinie ist besonders steil im Bereich wo beide Transistoren in Sättigung sind

Folie 43

Der Graph zeigt zwei Kennlinien für verschiedene W/L Faktoren

Folie 44

Wir sehen dass die Kennlinien stark nichtlinear sind.

Das ist bei einem Inverter erwünscht:

Wir möchten, dass für eine Eingangsspannung, die der logischer 1 entspricht, der Ausgang auf jeden Fall logische 0 ist.

Auch wenn der Eingang logische 0 ist, sollte der Ausgang logische 1 sein.

Folie 45

Wie definieren wir logische Niveaus?

Logische Eins könnte z.B. als $V > V_{DD} - V_{th}$ definiert werden.

Logische Null als $V < V_{th}$.

Wir können die Bereiche im Graph skizzieren.

Folie 46

Wir sehen, dass die Bedingungen Eingang 1 -> Ausgang 0, und andersrum, immer erfüllt sind. Darüber hinaus, gibt es einen Sicherheitsabstand. Nämlich für $V_{in} < V_{th}$ $V_{out} \sim V_{DD}$ also zur Grenze der logischen 1 gibt es einen Abstand von V_{th} .

Folie 47

Solchen Sicherheitsabstand gibt es nur wenn die Kennlinie genug nichtlinear ist.

Folie 48

Eine vollständig lineare Kennlinie würde keinen Sicherheitsabstand bieten – wie in Folie 48 dargestellt ist.

Folie 49

Wie sieht ein Inverter in Wirklichkeit aus?

Die MOSFETs haben die folgenden Kontakte: Source, Drain, Gate.

Der NMOS befindet sich in einem P-Typ Substrat, der PMOS in einem N-Typ Substrat.

Normalerweise wird ein Chip auf einem p-Substrat gemacht. Das heißt, für einen PMOS wird ein lokales n-Typ-Substrat gebraucht. Dieses lokale Substrat nennen wir n-Wanne. Manchmal befinden sich auch NMOS Transistoren in lokalen p-Wannen.

Ein 3-D Bild des CMOS Inverters wird in Folie 49 gezeigt.

Zusätzlich zu Source, Drain und Gate, haben wir auch die Substratkontakte. Das sind die, so genannten, ohmschen (Tunneleffekt-) Kontakte – p+ Diffusion in P-Substrat und n+ Diffusion in N-Wanne.

Beachten wir auch, dass es in allen Bereichen wo wir keine Diffusion-Lagen haben „Dellen“ befinden, die mit SiO₂ gefüllt sind (engl. Trench-Oxide oder das Feldoxid). Diese dienen zur Isolierung von Transistoren.

Das Bild in Folie 49 zeigt die Skizze der CMOS Struktur nach der Herstellung.

Folie 50

Die Abbildung in Folie 50 zeigt die technische Zeichnung – das Layout, die für die Fertigung der CMOS Struktur benutzt wird.

Im Layout werden alle Lagen, 2D-, übereinander gezeichnet. Es gibt dabei einige Besonderheiten. Nicht alle Lagen werden gezeichnet. Z.B., wir zeichnen nicht direkt N- und P-Diffusionsbereiche sondern nur die Bereiche mit dem dünnen- und dem dicken Oxid.

Die Bereiche innerhalb der Lage „Diffusion“ enthalten das dünne Oxid, die Bereiche draußen, das dicke Feldoxid.

Wenn ein Bereich mit dem dünnen Oxid noch mit der Lage „NPLUS“ umgeben wird, und wenn es darüber keine „Poly“ Lage befindet - wird aus dem Bereich ein n+ Diffusionskontakt. Das kann ein NMOS Source/Drain oder ein P-Substrat kontakt sein.

Der Überlapp von „Diffusion“-Lage, dem „Poly“ und dem „NPLUS“ bildet ein NMOS Gate.

Ein „CA“ (Kontakt) über einer „Poly“ – Lage ist ein Polysilizium – Metall 1 (M1) Kontakt.

Ein CA über einem Diffusionskontakt ist der Kontakt zwischen dem Silizium-Substrat und der ersten Metalllage.

Folie 51

Wenn wir einen Inverter als Teil einer Standardzellenbibliothek verwenden möchten, müssen wir dafür sorgen, dass die Synthese Tools solche Invertern leicht nebeneinander platzieren können. Das heißt z.B., die Verbindungen für VDD oder GND sollen übereinstimmen. Man darf normalerweise nur die erste Metalllage (M1) für die internen Verbindungen verwenden (VDD, GND, Drains, Gates), so dass die Standardzellen mit weiteren Metallagen verbunden werden können.

Es ist auch wichtig, dass sich Eingänge und Ausgänge an zugänglichen Stellen befinden, die in einer Textdatei definiert werden.

Generell gilt, die Höhe von allen Standardzellen ist gleich. Die Breite von Standardzellen ist ein Vielfach einer Breitereinheit. Das ermöglicht den Tools eine leichte Platzierung von den Standardzellen.

Folie 52

Diese Folie zeigt das Layout eines AND-Gates.

Folien 53-68

Versuchen wir die Geschwindigkeit eines Inverters zu berechnen:

Wir nehmen an, dass wir am Ausgang des Inverters eine kapazitive Last haben. (Folie 53) Die Annahme ist auch, dass wir am Eingang einen unendlich schnellen

Impuls haben – die Eingangsspannung ändert sich von GND auf VDD. In dem Fall wird NMOS momentan eingeschaltet und PMOS ausgeschaltet. (Folie 54)

Folie 55

Die Ausgangsspannung war am Anfang VDD. Wir verwenden die ID-VDS Kennlinie des NMOS Transistors.

Folie 56

Im Ausgangsbereich zwischen VDD und VDD – Vth (Bereich 1) wird der Kondensator mit konstantem Strom entladen.

Folie 57

Im weiteren Bereich 2 hängt der Entladestrom vom VDS ab.

Folien 58-59

Wir werden den ersten Teil aus folgenden Gründen vernachlässigen: Der Unterschied zwischen logischer 1 und 0 (logischer „Swing“) ist deutlich größer als Vth. Der Entladestrom im Bereich 1 ist am größten. Deshalb befindet sich das Potential VDS viel kürzere Zeit im Bereich 1 als im Bereich 2. Die Entladezeit im Bereich 2 dominiert.

Folie 60

Das Entladeverhalten im Bereich 2 kann mit folgender Differentialgleichung hergeleitet werden:

$$C \frac{dU}{dt} = - I_{ds}$$

Oder

$$C \frac{dU}{dt} = - k (V_{ds} V_{gst} - \frac{V_{ds}^2}{2}) = - k (V_{gs} U - \frac{U^2}{2})$$

$$V_{gst} = V_{gs} - V_{th}$$

Normalerweise sind nichtlineare Gleichungen schwer zu lösen – diese, aber, nicht. Sie kann analytisch gelöst werden

Die Variablen werden getrennt:

$$dU/(V_{gs} U - U^2/2) = -k/C dt$$

Beide Seiten werden integriert – die Gleichung gilt für $U < V_{gst}$

Die Lösung ist

$$U(t) = 2V_{gst} \exp(-t/\tau)/(1+\exp(-t/\tau))$$

τ ist hier $T = C/K$

Folie 61

Beachten wir, dass sich im Bereich um $V_{ds} = 0$ der Transistor wie ein Widerstand mit $R_{on} = 1/K$ verhält. Die Formel oben kann dann wie folgend umgeschrieben werden: (Wir setzen $R_{on} = 1/K$ ein.)

$$U(t) = 2 V_{gs} \exp(-t/R_{on}C)/(1 + \exp(-t/R_{on}C))$$

Diese Formel ist sehr ähnlich wie die Formel wenn wir eine Kapazität mit einem linearen Widerstand entladen würden:

$$U(t) = U(0) \exp(-t/R_{on}C)$$

Folie 62

Folie 62 zeigt den Unterschied zwischen zwei Funktionen. Im ersten Fall dauert ist die Entladezeit etwa $3 \cdot RC$ im zweiten $4 \cdot RC$.

Folie 63

Wichtig ist das Folgende – die Geschwindigkeit des Inverters, also die Entladezeit hängt direkt von der Lastkapazität ab und umgekehrt vom Faktor K. Faktor k hängt von der Mobilität der Ladungsträger (μ) und vom Verhältnis W/L ab.

Also $T \sim C / (\mu W/L)$

μ und W/L sind die Parameter vom NMOS

Im Fall wenn sich der Eingang von VDD auf GND ändert, gilt die gleiche Formel mit dem Unterschied dass wir μ und W/L vom PMOS haben.

Folie 64

Wenn fallende und steigende flanke gleich sein sollen, müssen wir unterschiedliche Mobilitäten für Elektronen und Löcher mit verschiedenen W/L Faktoren kompensieren.

Deshalb sind die PMOS Transistoren normalerweise breiter.

Folie 65

Welche Kapazität stellt C dar?

Das könnte z.B. die Kapazität von der digitalen Zelle sein, die an Inverter angeschlossen ist.

Folie 66

Oft werden Invertern als Treiber für die Takt-Leitung verwendet. Eine Taktleitung ist an viele Flip Flips angeschlossen und hat deshalb eine große Kapazität.

In diesem Fall brauchen wir Invertern mit großem W/L Verhältnis.

Folie 67

Das bedeutet in Realität, dass es in einer digitalen Bibliothek Invertern mit verschiedenen Stärken gibt. Ein kleiner Inverter hat die Stärke 1 (oder 0) und wird als INV_1 (oder als INV_0) bezeichnet. Es gibt dann größere Invertern INV_2 ... 4 ... 8.

Folie 68

Oft entspricht ein INV_2n zwei, in parallel, geschalteten INV_n. Das Layout ist normalerweise angepasst, so dass INV_2n nicht unbedingt im Layout 2x größer ist, obwohl sein effektives W/L Verhältnis 2x größer ist.

Inverter-Kette

(Keine Folien)

Die Wahl des Inverters wäre einfach wenn die Invertern selbst keine Kapazitive Last für die vorherige Stufe erzeugen würden. Wir könnten dann immer den größten Inverter nehmen.

Woher kommt die Kapazität eines Inverters? Erinnern wir uns, dass die MOSFET Transistoren eine Oxidschicht zwischen dem Gate und dem Kanal haben. Das heißt, es gibt eine Gate-Kapazität mit der Größe

$$C_{ox} = W * L * \text{EpsilonSiO2} / T_{ox}.$$

Eine detailliertere Analyse zeigt, dass die Gate-Kapazität auch vom Arbeitsbereich Abhängig ist – sie ist 2/3 C_{ox} in Sättigung, volle C_{ox} im Linearbereich, usw.

Ein n-facher Inverter hat also eine n-fache Eingangskapazität.

Ich werde es mit der folgenden Optimierungsaufgabe illustrieren.

Wir haben einen Flip-Flop Ausgang, der einem Inverter mit Stärke 1 entspricht.

Wir möchten, dass der Flip-Flop ein Taktsignal generiert, das für 1100 weitere Flip-Flops verwendet wird. Wir haben die Invertern mit Stärken 1, 2, 4 ... zur Verfügung. Die Frage ist, wie die optimale Lösung im Sinne der Taktsignal-Verzögerung aussieht.

Wir können z.B. Den Flip-Flop- (Q-) Ausgang direkt an die 1100 Flip-Flops anschließen, oder, eventuell, einen sehr starken (z.B. 128x-) Inverter dazwischen schalten, oder eine längere Kaskade von Invertern mit steigender Stärke benutzen.

(Ich vernachlässige hier die Tatsache, dass ein Inverter den Takt invertiert – es steht immer Q und QB (negiert) zur Verfügung.)

Wenn wir also Q an 1100 FFs anschließen, können wir erwarten, dass die Takt-Anstiegszeit zu langsam ist. Die Taktfrequenz ist dann begrenzt.

Wenn wir 128-fachen Inverter an QB anschließen, ist dieser wahrscheinlich schnell genug um Takteingänge von 1100 FFs zu treiben. Das Problem ist dann, dass ein 128-facher Inverter selbst eine große Kapazitive Last darstellt, die ungefähr den 128 Flip-Flops entspricht. Der QB-Ausgang im Flip-Flop (eigentlich auch ein Inverter) wird wegen der großen Lastkapazität stark verlangsamt. Es ist also vermutlich am besten, mehrere Invertern zu nacheinander schalten, z.B. INV_M, dann INV_N ($N > M$). Die Frage ist wie viele und wie sollen die Größenverhältnisse N, M, ... sein?

Es ist interessant, dass man diese mathematische Optimierungsaufgabe analytisch lösen kann.

Das Ergebnis ist sehr einfach. Wir brauchen eine Kaskade von Invertern. Optimal wäre, dass der nächste Inverter immer um Faktor $e = 2.718...$ größer ist als der vorherige. Um eine kapazitive Last zu „treiben“, die z.B. 1100 INV_1 Invertern entspricht, beginnend von einem INV_1 brauchen wir $\ln(1100) = 7$ Invertern:

Die Stärken sind: 1x, 2.7x, 7.9x, 20x, 55x, 148x, 402x.

Normalerweise wird statt $e=2.718...$, ein Verhältnis 2x oder 3x verwendet, da es im Layout einfacher zu realisieren ist.