

1. Transistorgleichungen

1.1. Bipolar

$$I_F = I_{ES} \left(\exp \left(\frac{U_{BE}}{U_T} \right) - 1 \right)$$

Vorteil: bessere Treiberverstärkung durch Exponentialfunktion Nachteile:

- hohe Verlustleistung
- langames Abschalten durch Sättigung mit Ladungsträgern → ECL-Schaltung / Differentielles ECL / Feedback-ECL Differenzverstärker

1.2. MOS-Transistoren

$$\text{Verstärkung: } \beta = K' \frac{W}{L} \text{ mit } K' = \frac{\mu \varepsilon_{ox} \varepsilon_0}{t_{ox}} \quad [\beta] = \frac{A}{V^2}$$

Kanalweite	W
Kanallänge	L
Elektronenbeweglichkeit	$\mu_n \approx 250 \cdot 10^{-4} \frac{m^2}{Vs}$
	$\mu_p \approx 100 \cdot 10^{-4} \frac{m^2}{Vs}$
rel. Dielektrizität des Gateoxys	$\varepsilon_{ox} \approx 3,9$
Dielektrizitätskonstante	$\varepsilon_0 = 8.8541878 \cdot 10^{-12} \frac{As}{Vm}$
Gateoxydicke	t_{ox}
Verstärkung	$\beta = \frac{\mu_n \varepsilon_{ox} \varepsilon_0}{t_{ox}} \cdot \frac{W}{L} = K' \frac{W}{L}$
Kapazität	$C_G = \varepsilon_{ox} \varepsilon_0 \frac{WL}{t_{ox}}$
Verzögerungszeit	$t_{pHL} \propto \frac{C_L t_{ox} L_p}{W_p \mu_p \varepsilon_{ox} (V_{DD} - V_{th})}$

- große Kanalweite ⇒ große Drain-Störme
⇒ schnelle Schaltgeschwindigkeit (da $i_d \propto \beta \propto \frac{W}{L}$)
Aber: große Fläche.
- nMos schaltet schneller als pMos, da nMos und pMos unterschiedliche Majoritätsladungsträger haben. Die Beweglichkeit der Löcher ist im Allgemeinen geringer als die der Elektronen.
- nMos zum Entladen, pMos zum Aufladen

nMos (p-dotiertes Substrat, n-dotierte Drain/Source), schlechter pull up (Pegeldegenerierung)

$$I_d = \begin{cases} 0, & \text{für } U_{gs} - U_{th} \leq 0 \\ \beta [(u_{gs} - U_{th}) \cdot u_{ds} - \frac{1}{2} u_{ds}^2], & \text{für } 0 \leq U_{gs} - U_{th} \leq u_{ds} \\ \frac{1}{2} \beta \cdot (u_{gs} - U_{th})^2, & \text{für } 0 \leq U_{gs} - U_{th} \leq u_{ds} \end{cases}$$

pMos (n-dotiertes Substrat, p-dotierte Drain/Source), schlechter pull down (Pegeldegenerierung)

$$I_d = \begin{cases} 0, & \text{für } U_{gs} - U_{th} \geq 0 \\ -\beta [(u_{gs} - U_{th}) \cdot u_{ds} - \frac{1}{2} u_{ds}^2], & \text{für } 0 \geq U_{gs} - U_{th} \geq u_{ds} \\ -\frac{1}{2} \beta \cdot (u_{gs} - U_{th})^2, & \text{für } 0 \geq U_{gs} - U_{th} \geq u_{ds} \end{cases}$$

2. Sequentielle Logik

t_{Setup}	Stabilitätszeit vor der aktiven Taktflanke
t_{hold}	Stabilitätszeit nach der aktiven Taktflanke
t_{c2q}	Eingang spätestens nach t_{c2q} am Ausgang verfügbar
Min. Taktperiode	$t_{clk} \geq t_1, c2q + t_{logic, max} + t_{2, setup}$ (bei Verletzung: neue Pipelinestufe aber dadurch höhere Latenz)
Max. Taktfrequenz	$f_{max} = \left\lfloor \frac{1}{t_{clk}} \right\rfloor$ (Nicht aufrunden)
Holdzeitbedingung	$t_{hold} \leq t_{c2q} + t_{logic, min}$ (bei Verletzung zwei Inverter einfügen)
Durchsatz	$\frac{1 \text{ Sample}}{t_{clk, pipe}} = f$
Latenz	$t_{clk} \cdot \# \text{Pipelinestufen}$ (das zwischen den FFs)

2.1. Zeitmessungen

$$\text{Ersatzwiderstand für den Transistor: } R_{on} \approx \frac{1}{\beta (|U_{gs, p}| - |U_{th, p}|)}$$

$$\text{Guardband: } t_{guard} = t_{clk} - t_{p, max}$$

2.1.1. Propagation-Delay

Zeit, die beim Umschalten zwischen 50% der Eingangsspannung und 50% der Ausgangsspannung vergeht

$$t_p = \frac{t_{p, LH} + t_{p, HL}}{2} = \ln 2 R_{on} C_L$$

Für eine Treiberstufe: $t_p = \ln 2 \frac{R_{on}}{S} (S \cdot C_{int} + C_L)$ mit Kanalweiten-

$$\text{skalierung } S = \sqrt{\frac{C_L}{C_g}}$$

$$t_{p, LH} \approx \frac{C_L t_{ox} L}{W \mu \varepsilon (U_{DD} - |U_{th, p}|)} \text{ Falls } W \rightarrow \infty \text{ gilt trotzdem } t \neq 0, \text{ da interne Gate-Kapazität so wächst!}$$

2.1.2. Rise- and Fall-Time

Zeit, die beim Umschalten eines Signals von 10% auf 90% vergeht

2.1.3. Lastkapazitäten

Umladen von parasitären Lastkapazitäten braucht Zeit. Diese setzen sich wie folgt zusammen:

- Eingangs/Gate-Kapazität der nachfolgenden Stufe
- Drain-Kapazität
- Leiterbahn-Kapazitäten

2.1.4. Elmore-Delay

Es müssen alle Kapazitäten berücksichtigt werden und alle Widerstände auf gemeinsamen Pfaden. Es sollen dann die RC Konstanten aufaddiert werden.

$$\text{Beispiel: } \tau = R_C \cdot C_2 + (R_C + R_B) C_1 + (R_C + R_B + R_A) C_L$$

2.2. Pipelining

Nur bei synchronen (taktgesteuerten) Schaltungen möglich!

- Aufteilen langer kombinatorischer Pfade durch Einfügen zusätzlicher Registerstufen
→ Möglichst Halbierung des längsten Pfades
- Zeitverhalten beachten (evtl. Dummy-Gatter einfügen)
- Durchsatz erhöht sich entsprechend der Steigerung der Taktfrequenz
- Gesamtlatenz wird eher größer
- Taktfrequenz erhöht sich

2.3. Latch

SB = RB = Q = QB ⇒ Latch schwingt

$t_{rise} \neq t_{fall} \Rightarrow$ Schwingung klingt ab

3. Statische CMOS-Schaltungen

XXX

4. Pass-Transistor-Logik

Vorteile	Nachteile
Wenige Transistoren	zusätzliche Inverter oder Transmission Gates für Pegelregeneration
Kleiner Flächenverbrauch	Verdrahtungsaufwand hoch
Geringe Verlustleistung	Langsam

4.1. Erweitertes KV-Diagramm

Einträge (0 oder 1) des KV-Diagramms werden durch die jeweilige DNF für den Einseintrag ergänzt. Hierdurch können anschließend einfach die Einträge mit den gleichen DNFs gefunden werden.

5. Dynamische CMOS-Schaltungen

Vorteile	Nachteile
Nur ein Logikblock statt zwei benötigt	Ladungsverlust durch Leckströme und und Charge-Sharing

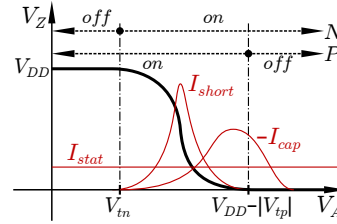
Damit die Pegelregeneration gewährleistet ist, sollten Gatter immer abwechselnd mit NMOS und PMOS-Technik kaskadiert werden. Alternativ kann auch eine statische Schaltung benutzt werden.

5.1. Master-Slave-Flip-Flop

Achtung mit Clock-Gating bei dynamischen Registern, sonst Datenverlust durch Leckströme

6. CMOS Verlustleistung

Inverterschaltvorgang $V_A : 0 \rightarrow 1$:



6.1. Dynamische Verlustleistung

$$P_{dyn} = P_{cap} + P_{short}$$

Kapazitive Verluste

$$P_{cap} = \alpha_0 f C_L V_{DD}^2$$

Kurzschlussstrom

$$P_{short} = \alpha_0 f \beta_n \tau (V_{DD} - 2V_{tn})^3$$

Schalzhäufigkeit

$$\alpha_0 \rightarrow 1 = \frac{\text{Schaltvorgänge (pos. Flanke)}}{\# \text{ Betrachtete Takte}}$$

Schalzhäufigkeit (periodisch)

$$\alpha = \frac{f_{clk}}{f_{clk}}$$

Abhängig von den Signalfanken, mit Schaltfunktionen verknüpft

$$\approx V_{DD} 1 / \propto \text{Schaltzeit: } \frac{V_{DD} 2}{V_{DD} 1} = \frac{t_{D1}}{t_{D2}} \text{ (bei Schaltnetzen } t_{log})$$

$$\text{Verzögerungszeit} \propto \frac{C_L t_{ox} L_p}{W_p \mu_p \varepsilon (V_{DD} - V_{th})}$$

Steigend mit: Kapazitiver Last, Oxiddicke, Kanallänge, Schwellspannung
Sinkend mit: Kanalweite, Ladungsträger Beweglichkeit, Oxyd Dielektrizität, Versorgungsspannung

6.2. Statische Verlustleistung

Abhängigkeit: $V_{DD} \uparrow: P_{stat} \uparrow \quad V_{th} \uparrow: P_{stat} \downarrow$ (aber nicht proportional)

Sub-Schwellströme:

$$I_D = I_0 \exp \left(\frac{V_{gs} - V_T}{n V_{Temp}} \right) (1 - \exp (V_{ds} - V_{Temp})) \text{ für } V_{gs} < V_T$$

Leckströme/Gate-Ströme (Sperrströme, da Gate-Oxit nicht richtig isoliert):

$$I_{Gate} \sim \exp \left(t_{ox}^{-1} \right)$$

Optimierung durch:

- Clock-Gating: Nicht benötigte Blöcke können innerhalb einer Taktflanke an- und abgeschaltet werden
- Power-Gating: Nicht benötigte Bereiche können abgeschaltet werden, Daten müssen jedoch zuvor abgespeichert werden
- Dynamic Voltage Frequency Scaling: Versorgungsspannung und Frequenz $\downarrow \Rightarrow$ Dynamic Power \downarrow

Die Schwellspannung kann durch Substratvorspannung (Bulk) verändert werden

7. Speicher

7.1. DRAM-Zelle

Information wird nur temporär in Kondensator gespeichert. Es muss regelmäßig aufgefrischt werden.

7.2. SRAM-Zelle

TODO

7.3. Flash-Zelle

Elektronen werden durch eine hohe Spannung am Gate und Drain auf das Floating-Gate gezogen. Zur Speicherung wird keine Spannung benötigt. Beim Lesen mit mittlerer Spannung verschieben die Elektronen im Floating-Gate die Thresholdspannung. Das Oxid wird beim Laden/Entladen beschädigt (geringe Lebenszeit), deswegen muss die Nutzung auf allen Speicherzellen verteilt werden.

7.4. Phase-Change-Memory

Bei Raumtemperatur kann das Phase-Change-Material sich bei Raumtemperatur stabil in zwei Zuständen befinden: Amorph und Kristallin. Die Zustände können durch eine Reset(Melt)temperatur bzw. der Crystaltemperatur gewechselt werden.

8. Fehlertoleranz

Es können verschiedene Fehler entstehen:

- Manufacturing Variability (V_T)
- Soft-Errors (Fehler, die auftreten und wieder verschwinden)
- Alterung und Transistorperformanz (langsamer werdend)

Fehlermodelle:

a) Permanente Fehler: Stuck-At(0/1), Stuck-Open, Bridge-Fault (Ursache: Elektromigration)

b) Transiente Fehler: Kurzfristig falsche Werte, Single Event Transient/Upset (Ursache: natürliche Strahlung bsplw. alpha-Strahlen)

Single Event Transient: kurzer falscher Spannungswert durch Zeitverletzung durch Nebensprechen oder Alterung

Single Event Upset: Übernahme eines falschen Wertes durch Umkippen in ein FlipFlop

8.1. Fehlermaskierung

Fehlersensitivität: Fehler ist am Ausgang bemerkbar

Logikmaskierung: Wert ist irrelevant für den Ausgang

Transiente Maskierung: Logische Maskierung an späterer Stufe verhindert falsche Ausbreitung

9. Synchrone und Asynchrone Schaltungen

9.1. Synchrone Schaltungen

Taktverteilung meistens durch H-Baumstruktur, um alle Register in gleicher Zeit zu erreichen.

Taktabweichungen durch:

- Clock-Generation
- Devices
- Interconnect
- Temperatur
- Kapazitive Lasten
- Koppelkapazität zu benachbarten Leitungen

Clock-Skew	Clock-Jitter
Örtlicher Unterschied zw. Regionen	Zeitliche Unterschiede zw. aufeinanderfolgenden Taktflanken
Einsatz zur Verteilung der Stromlast auf größeren Zeitraum ⇒ Verkleinerung des IR-Drops	
Ausgleich von Unterschieden bei Pipelinestufen	

9.2. Asynchrone Schaltungen

Wie kann man einen sinnvollen Zeitpunkt finden um das Ergebnis aus einem asynchronen Zweig in ein Register zu übernehmen?

- Delay-Modul mit fester Zeit (start/done)

- Handshake-Protokoll mit redundanter Codierung (aber ständiger Querstrom!)
- Handshake-Protokoll mit Stromsensor (falls kein Strom mehr fließt → Schaltung fertig)

9.3. Kombination von asynchronen und synchronen Schaltungen

Synchrone Schaltung	Asynchrone Schaltung
+ einfache Timing-Analyse	+ schneller
+ einfache Verifikation	+ keine glob. Taktverteilung
- glob. Taktverteilung	- schwierige Verifikation
	- keine Timing-Analyse

Synchronisierung durch:

- asynchroner FIFO-Speicher mit Füllstandsüberwachung
- Handshake-Protokoll (Request/Acknowledgement)

Bei großen Schaltungen wird meistens die Globally-Asynchron-Locally-Synchron (GALS) Methodik benutzt. Hier werden synchrone Inseln mit asynchroner Technik verbunden.

10. Dynamische Rekonfiguration

Meist auf Field Programmable Gate Array-Chips (FPGA) realisiert und für Hardware-Beschleuniger benutzt.

tightly coupled (in CPU-Pfad) \Leftrightarrow loosely coupled (außerhalb)

Lookuptable (LUT): Speichern Ausgangswerte unabhängig von Eingangswerten

Configurable Logic Block (CLB): Block von der die Logik rekonfiguriert werden kann

Nachteile:

- Großer Overhead
- Große kapazitive Last und somit hohe Verlustleistung und geringere maximale Taktfrequenz

Das Ergebnis der Timinganalyse kann sich zwischen Pre- und Post-Layout durch Verdrahtung etc. unterscheiden.

Bei Timingverletzungen müssen zusätzliche Pipelinestufen eingefügt werden (evtl. durch Hyperregister in Switch-Box)

Pipelining in Rückkopplungsschleifen:

Partiell rekonfigurierbare Region (PRR) kann mehrere rekonfigurierbare Module (RM) enthalten

Alle statischen Leitungen in PRR müssen in allen RM vorkommen!