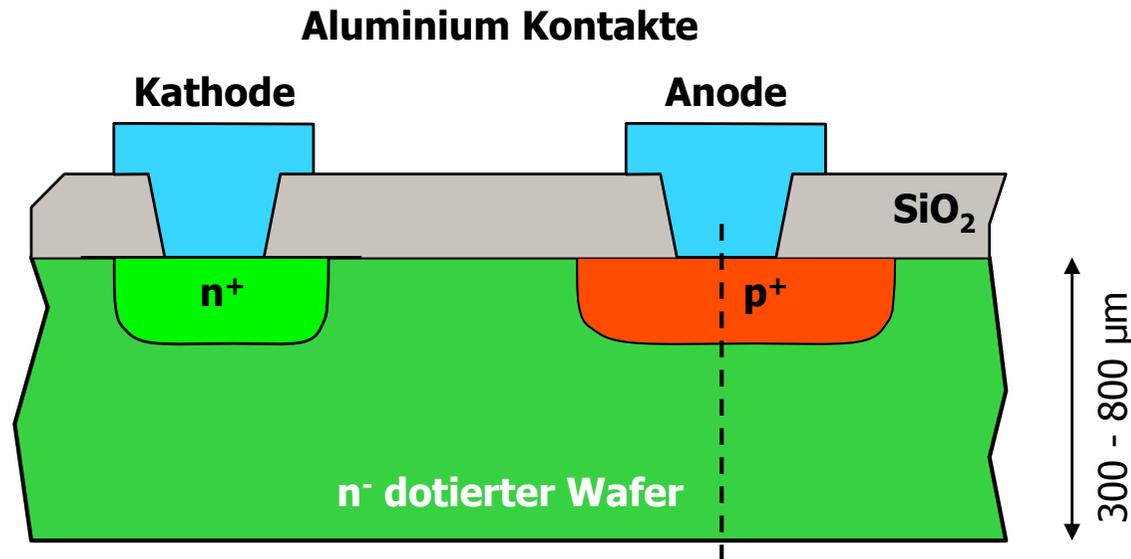

Layout von Bauelementen

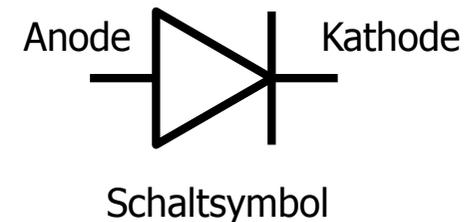
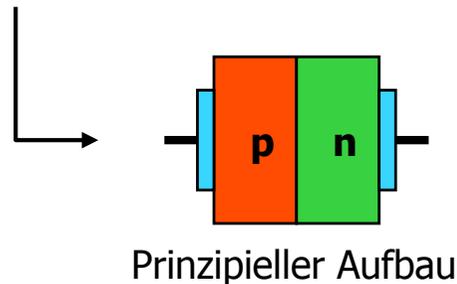
- MOS Transistoren
- Widerstände
- Kondensatoren
- Spulen
- Bipolare Transistoren in CMOS Prozessen
- Fuses (Sicherungen)
- Matching
- Verschiedenes: Elektromigration, Latchup..

pn-Diode durch Implantation

- In eine (z.B.) n- dotierte Si-Scheibe ('Wafer') werden an der Oberfläche stark dotierte Gebiete erzeugt
- JEDER pn-Übergang bildet eine Diode.
- Sie ist meist 'unerwünscht' – ein 'parasitäres' Element (insbesondere Drain, Source und Bulk des MOS)

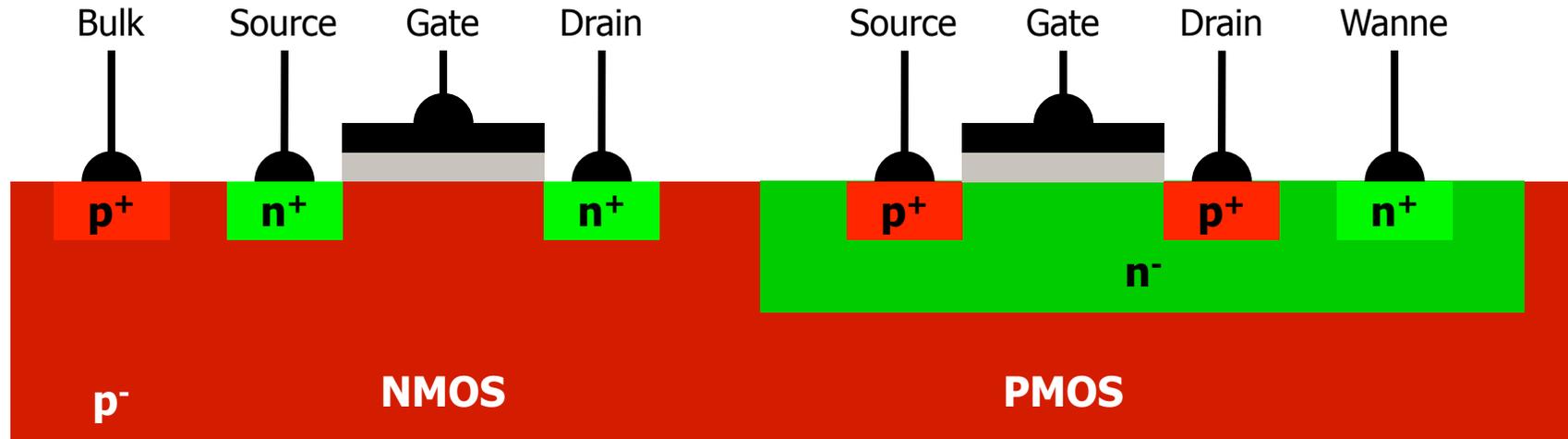


Querschnitt eines pn-Übergangs auf einem Chip

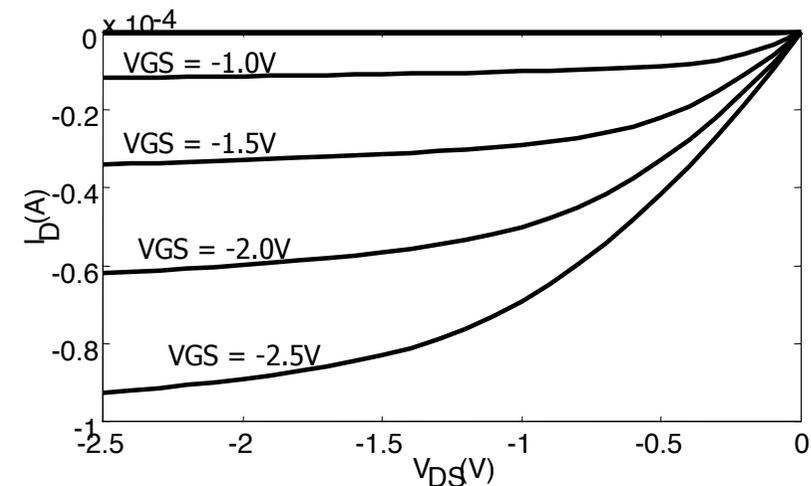


NMOS / PMOS

- Der **PMOS muss in einem n-Substrat sitzen**.
- Wenn der Wafer (für den NMOS) p-dotiert ist, wird eine n-dotierte **n-Wanne** (engl. 'well') eingebaut



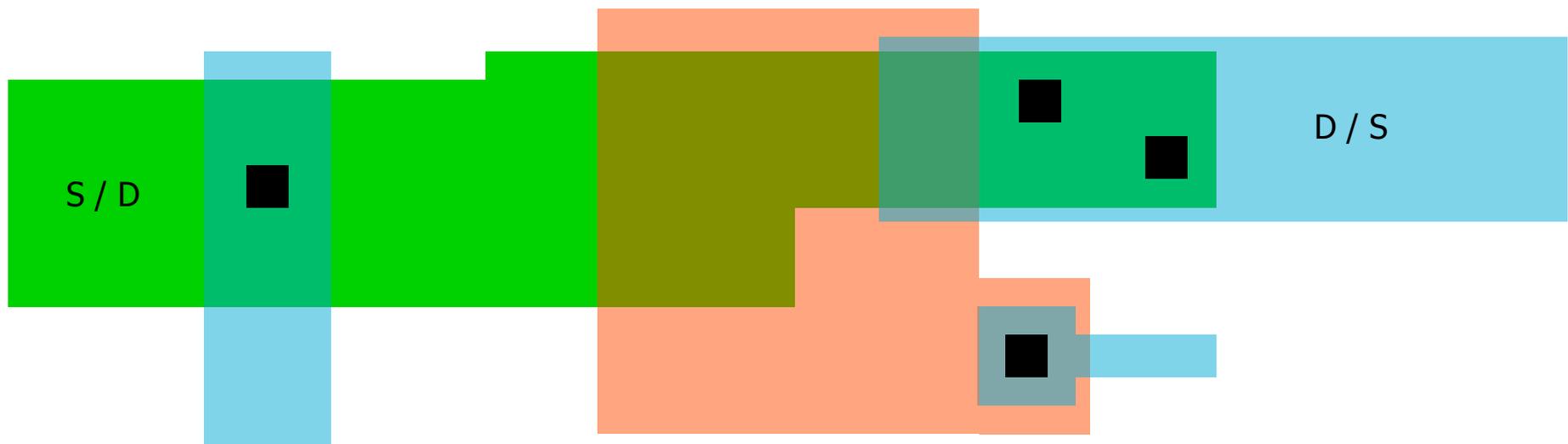
- Die Kennlinien sind genauso wie beim NMOS mit
 - negativen Polaritäten
 - **kleinerem K_p** (kleinere Mobilität der Löcher)



Transistoren

Transistoren

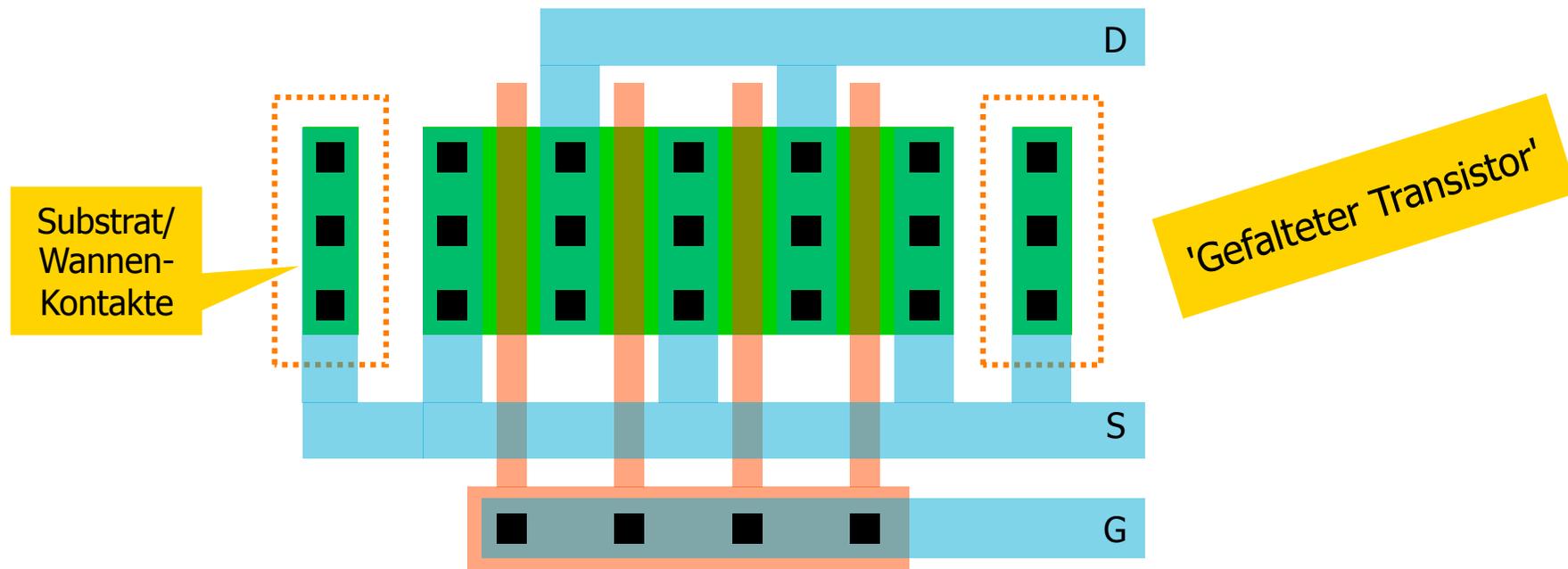
- Große Palette von W/L Verhältnissen
- Beim Layout beachten:
 - Zuleitungswiderstände zu Drain / Source / Gate
 - Kapazitätsverhältnisse
 - Matching zwischen identischen Bauteilen
- **Wannenkontakte** sind für JEDE Wanne notwendig. Meist wird die Wanne an die positive Versorgung angeschlossen. In analogen Layouts manchmal nicht (Beispiel: Sourcefolger). Man spricht dann manchmal von einem '**hot NWELL**'. Es gelten dann u.U. andere Abstandsregeln
- Das Substrat ist meist recht niederohmig, so daß wenige **Substratkontakte** reichen. Um **Latchup** (s. später) zu vermeiden, sollten jedoch so viele Kontakte wie möglich im Layout verteilt werden. Meist gibt der Hersteller eine Mindestdichte (Abstand des nächsten Kontakts von einem MOS) vor.
- Beispiel für ein schlechtes Transistorlayout:



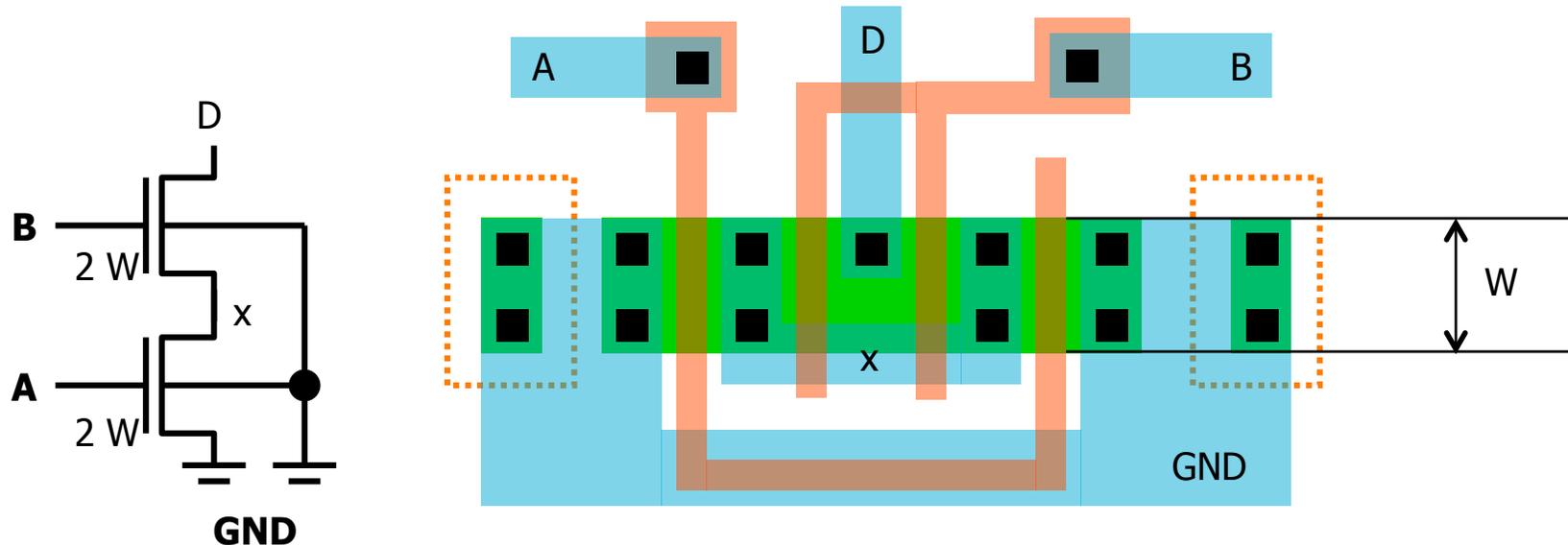
weite Transistoren (W groß)



- Bei großem W werden die Transistoren immer 'gefaltet':
 - + **halbe Drain/Source-Kapazität bei gleichem W !!!** (gerade Anzahl Teiltransistoren ist besser !)
 - + kleinerer Zuleitungswiderstand des Gates
- Es werden viele Drain/Source-Kontakte gezeichnet, um den Übergangswiderstand zu reduzieren



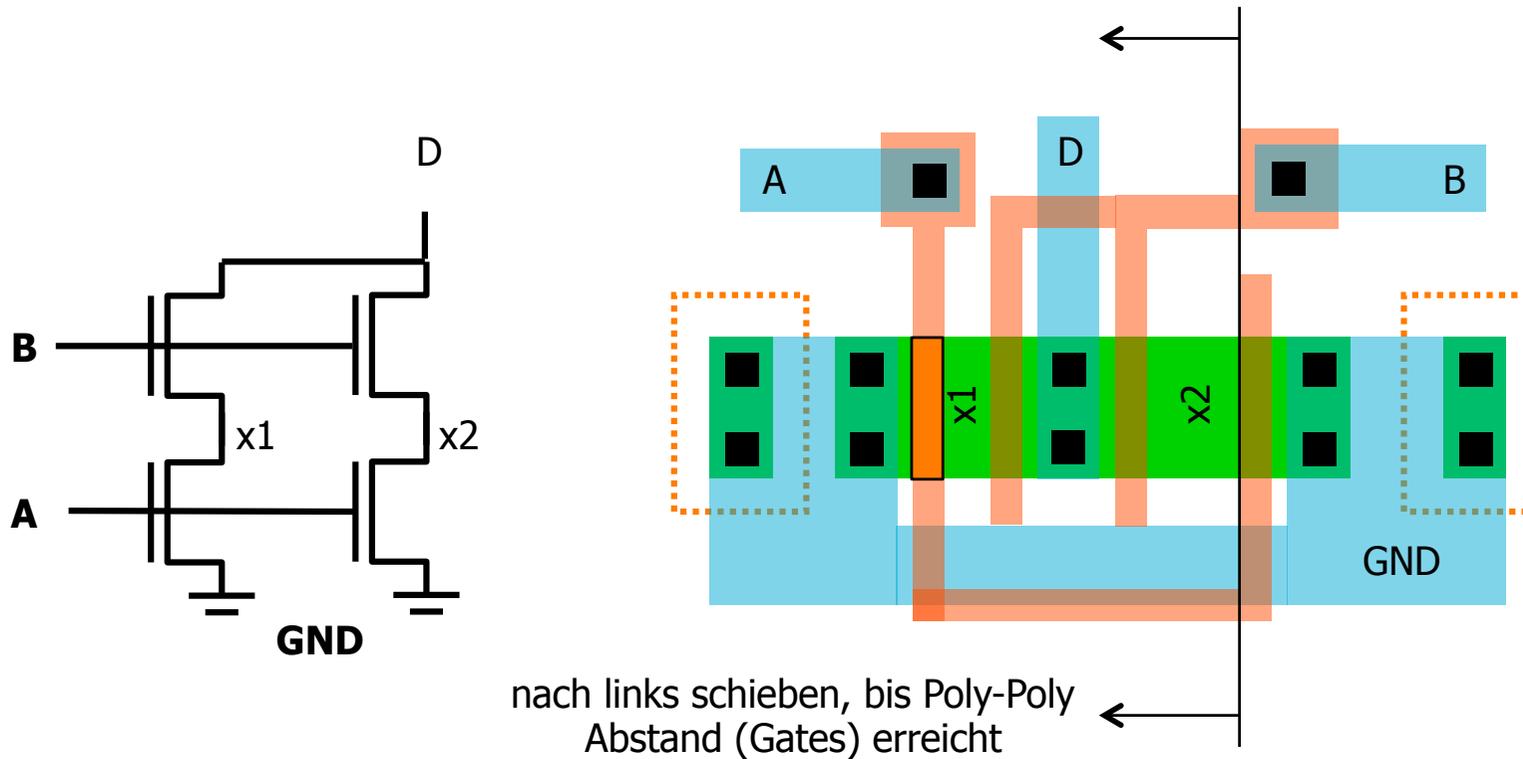
Beispiel: Zwei in Serie geschaltete weite MOS



- Hier: zwei NMOS in Serie wie z.B. beim NAND2 Gatter
- Substratkontakte hier sehr großzügig
- Es wäre dumm, den Ausgang außen und GND innen anzuschließen!

- Die Zwischenverbindung (x) kann man eigentlich weglassen, man benötigt dann auch keine Kontakte mehr → s. nächste Seite

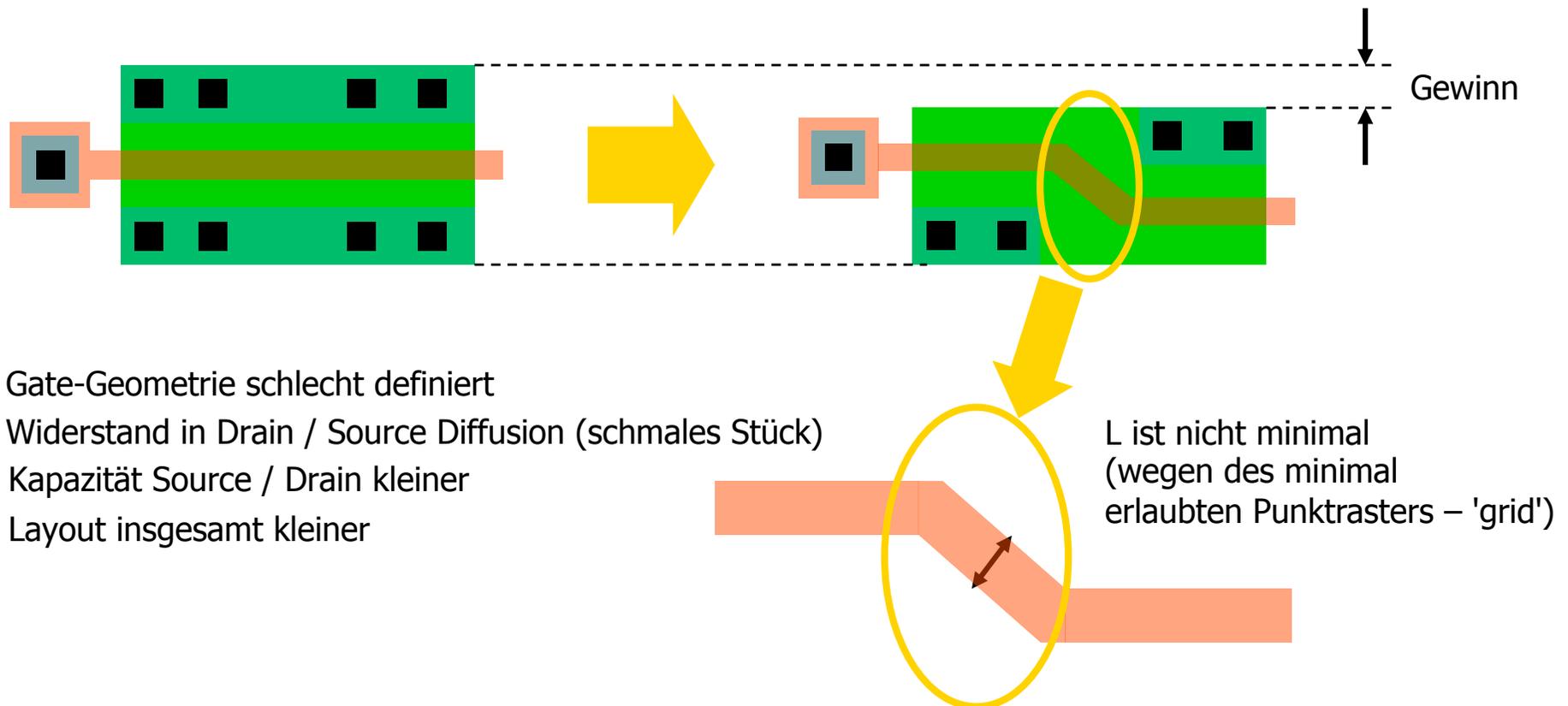
Optimierung



- Es gibt nun zwei unterschiedliche Knoten $x1/x2$. Diese müssen im Schematic getrennt sein!
- Die Kapazitäten dieser Knoten sind kleiner → gut!
- Da der $x1$ -Bügel nicht benötigt wird, kann die Masse weiter hoch rücken

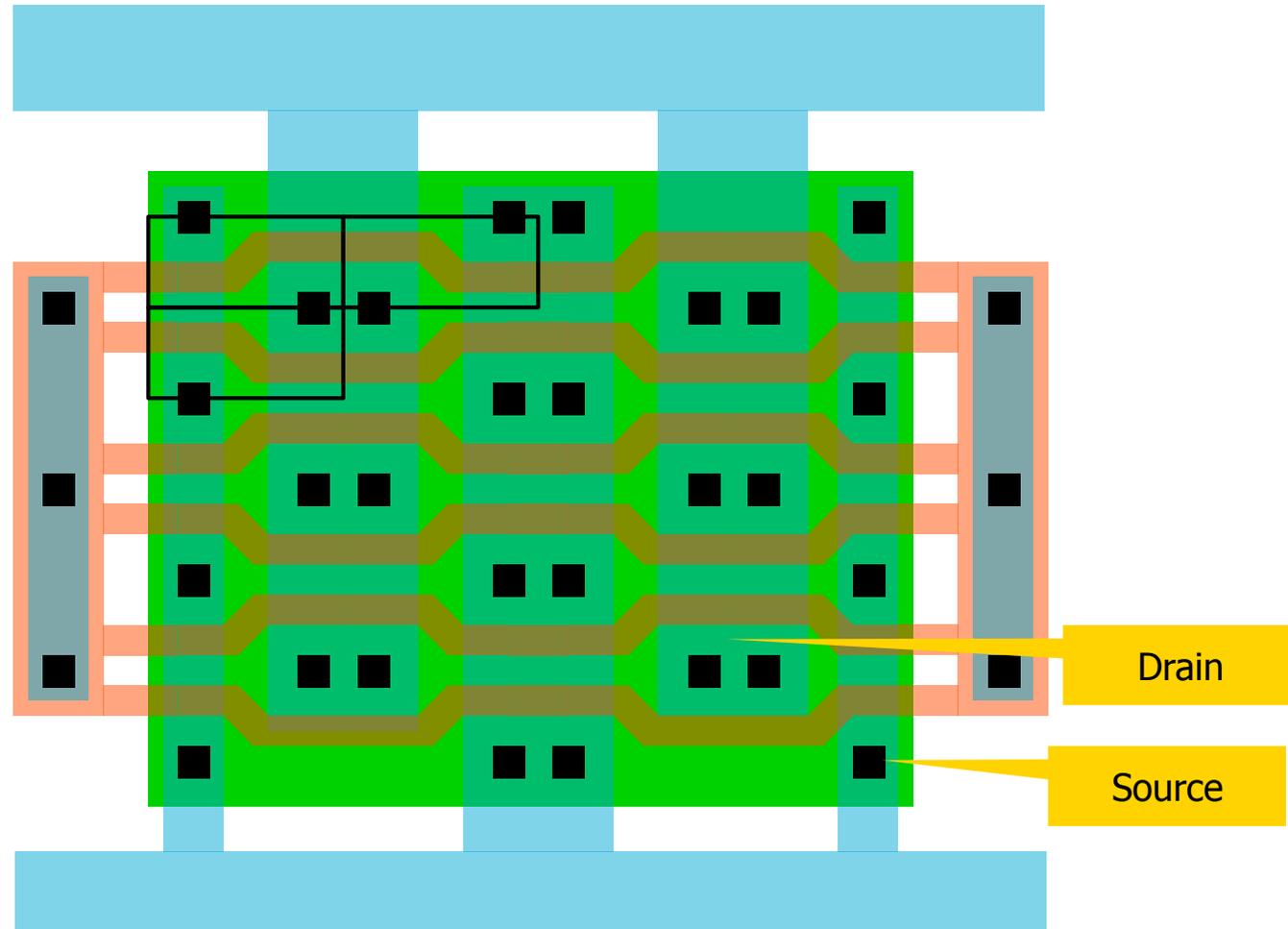
'Bent Gate' Transistoren

- Das Layout wird manchmal kompakter, wenn das Gate einen Knick ('bend') macht.
- W/L ist nicht ganz eindeutig \Rightarrow Messungen sind erforderlich



Große 'Bent Gate' Transistoren

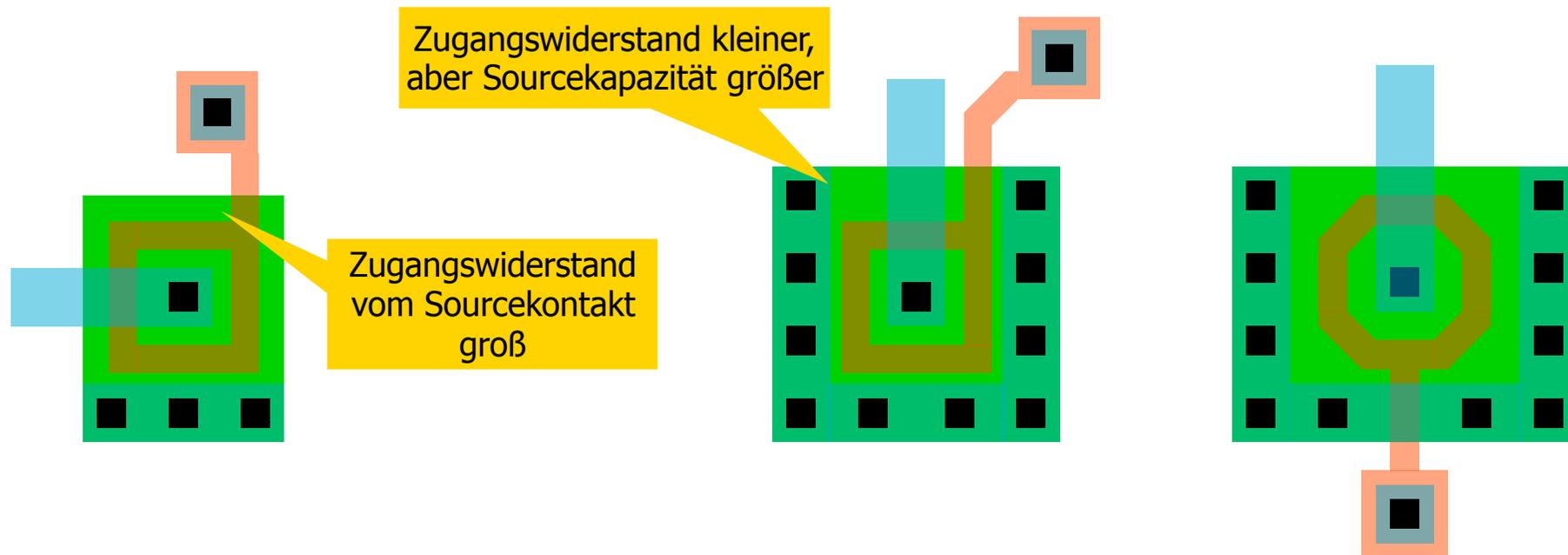
- Geometrie für 'viel W/L pro Fläche' bei gleichzeitig kleinen Zuleitungswiderständen
- ‚Waffle‘ Layout



- 'The Art of Analog Layout', p. 417

'runde' Transistoren

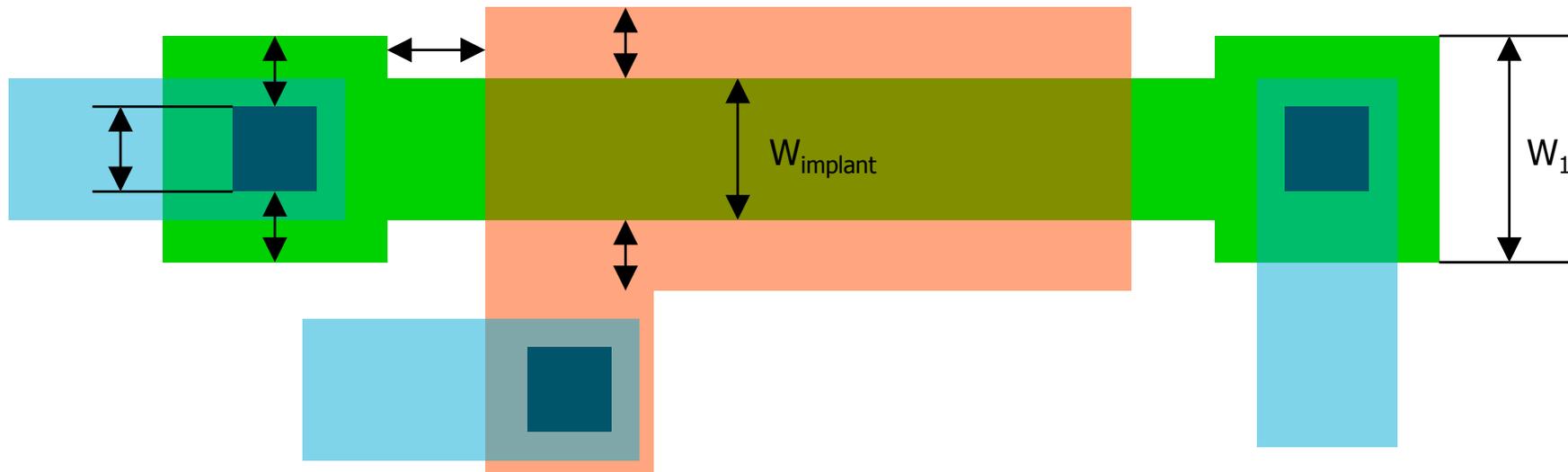
- Das kleinste (günstigste) Verhältnis von C_{GD} / g_m bekäme man mit **runden** Transistoren.
- Da diese meist die Design-Regeln verletzen (nur Vielfache von 45° erlaubt) zeichnet man **Rechtecke** oder **Achtecke**.
- Die Kapazitäten sind hier **SEHR** ungleich: $C_S \gg C_D, C_{GS} \gg C_{GD}$
- Nützlich z.B. für Ausgangs-Pads oder Open-Drain-Netze



- Das 'effektive' W/L dieser Strukturen ist schwer zu ermitteln. Messungen an Teststrukturen sind nötig!
- Die Extraktionsprogramme scheitern hier oft völlig (unsinnige L-Werte)

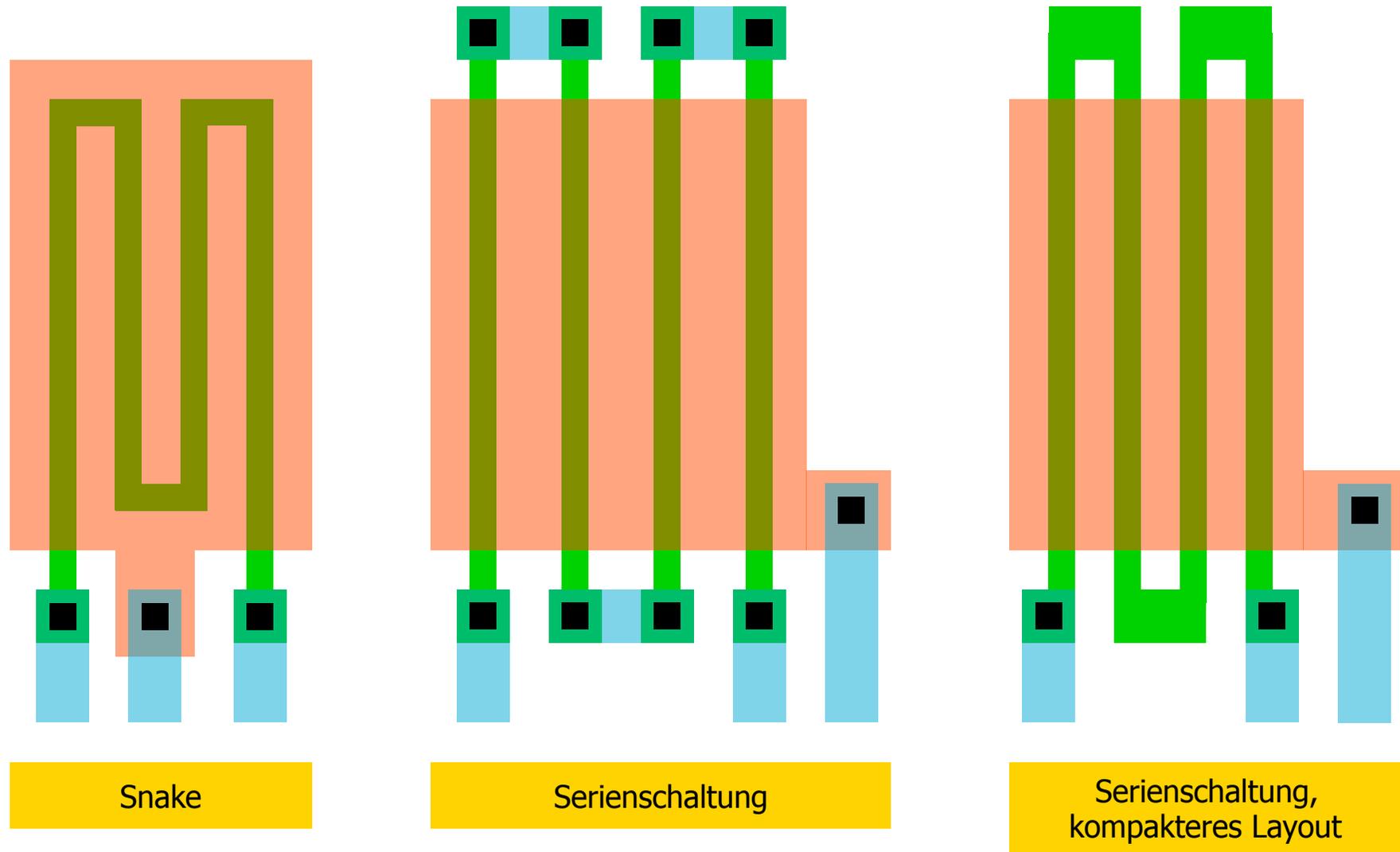
lange Transistoren (L groß)

- Minimales W der Implantation am Kontakt meist definiert durch $W_1 = W_{\text{Kontakt}} + 2 \times W_{\text{Überlapp}}$



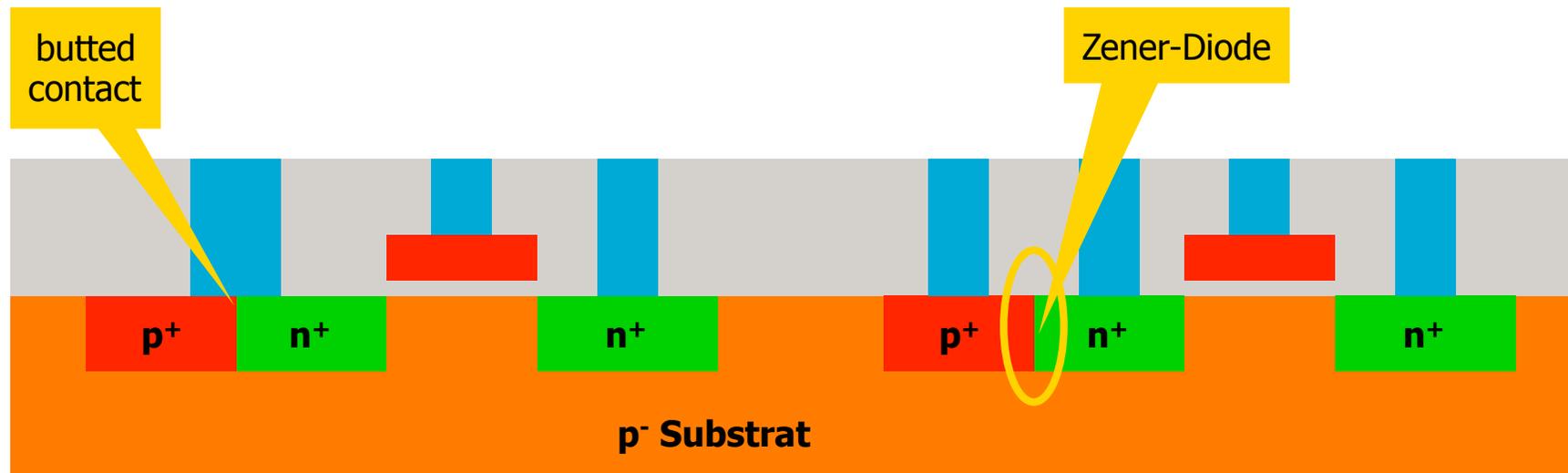
Sehr lange Transistoren (L groß)

- Layout meist als 'Snake'. Besser definierte Länge bei Serienschaltung von mehreren linearen MOS.



Butted Contacts

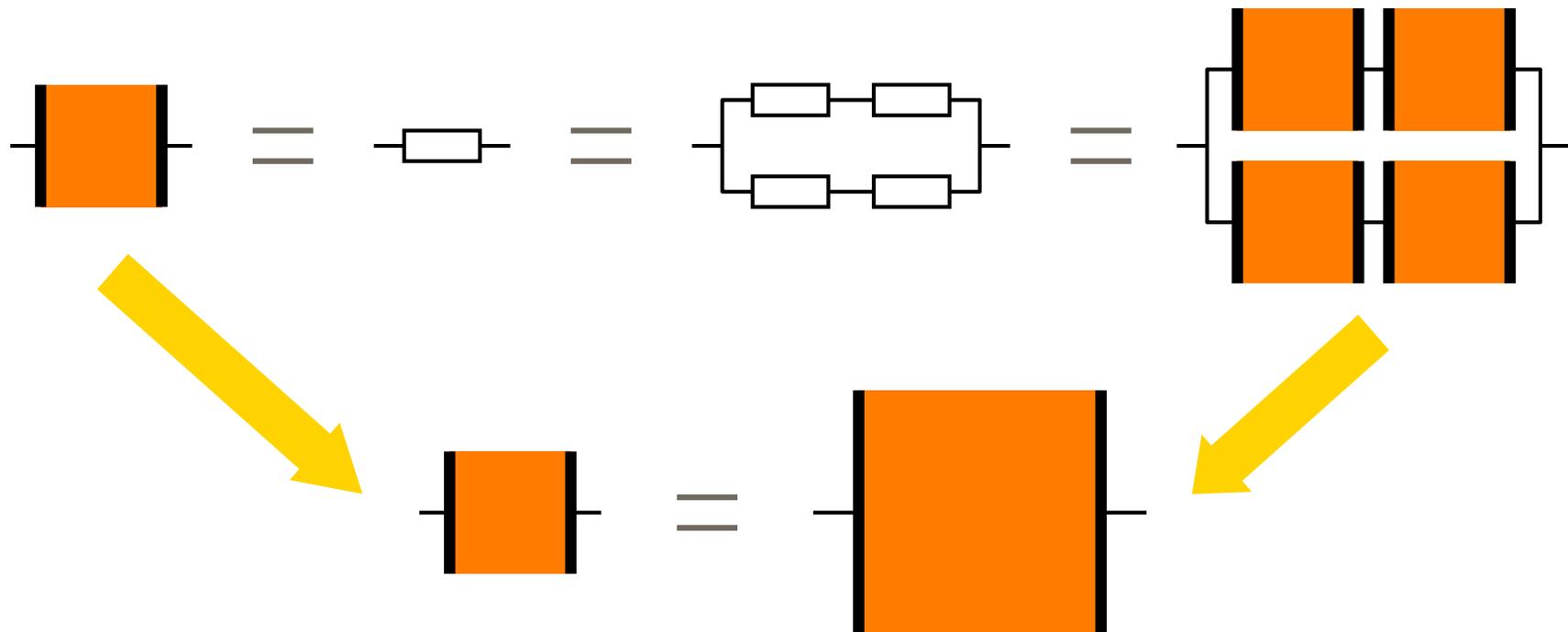
- Um kompakte Substratkontakte zu erlauben kann man einen gemeinsamen Kontakt über n^+ und p^+ legen.
- Man spricht dann von 'butted contacts'
- Sie sind in vielen Technologien nicht erlaubt
- Wenn sich n^+ und p^+ berühren, bildet sich u.U. eine Zenerdiode (hohe Dotierungen!). Solange beide Seiten auf gleichem Potential sind, ist das nicht schlimm, die Diode ist kurzgeschlossen.
- Für 'analoge' Transistoren ist es generell nicht gut, wenn sich n^+ und p^+ berühren!



Widerstände

Widerstände

- Alle Lagen (NWELL, N+, P+, Poly, Metalle) können für Widerstände benutzt werden
- Wichtigster Parameter ist der **Flächenwiderstand R (sheet resistance)**
- Seine Einheit ist '**Ohm per square**', R_{\square}
- Eine Schicht der Dicke t mit dem spezifischen Widerstand ρ [Ωcm] hat den Flächenwiderstand $R = \rho/t$
- Ein Quadrat hat immer den gleichen Widerstand, unabhängig von seiner Größe!



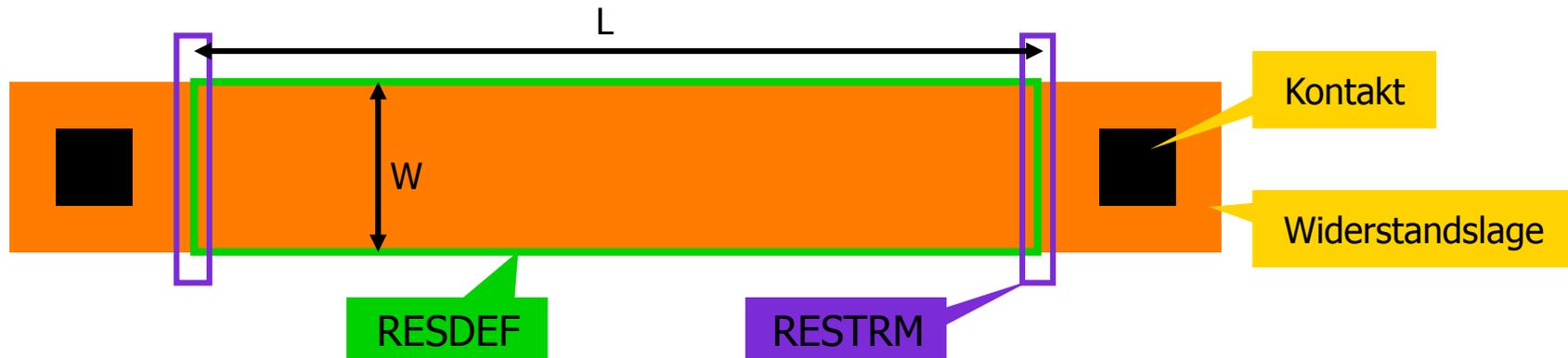
- Andere Betrachtung: Der Widerstand eines Quaders der Länge L , Breite W und Höhe t ist $R = \rho L/Wt$ also ist der Widerstand eines Quadrats ($W=L$): $R_{\square} = \rho/t = R$ unabhängig von der Größe

Rechteckige Widerstände

- Ein rechteckiger Widerstand hat den Wert

$$R = L/W \cdot R_{\square}$$

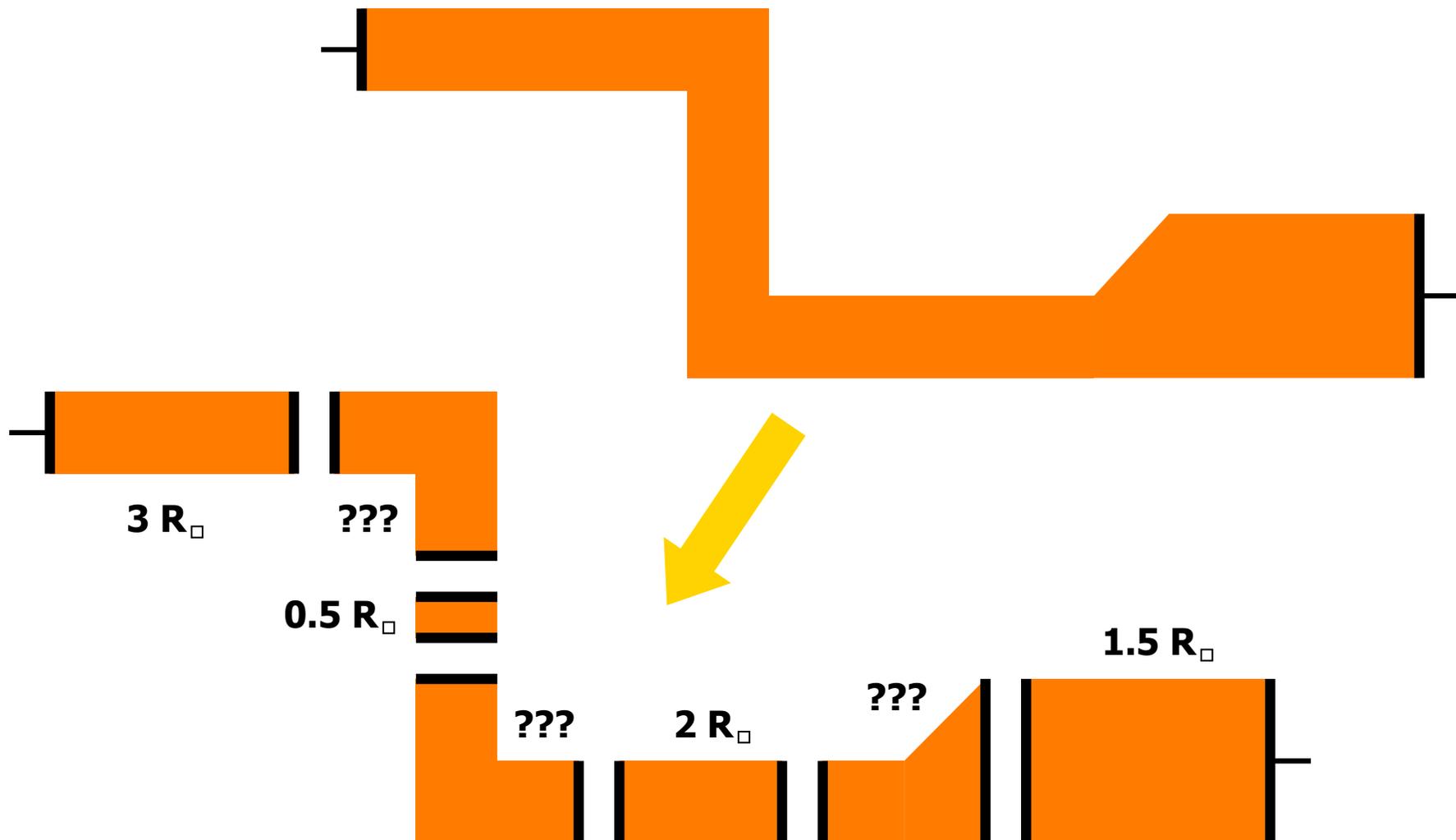
- Problem: Was sind W und L beim folgenden Layout ?



- Um der Extraktion zu 'helfen' werden der Widerstand und die Anschlüsse mit Sonderlagen markiert (in AMS Technologie: ‚RESDEF‘ und ‚RESTRM‘, in UMC: nur ‚PSYMBOL‘ anstelle ‚RESDEF‘)
- Details der Extraktion hängen vom Design-Kit der Technologie ab
- Strukturen ohne diese Lagen werden NICHT als Widerstände extrahiert!
- Der kurze Weg bis zu den Kontakten und die Kontakte selbst tragen auch einen kleinen Widerstand bei. Dieser ist meist vernachlässigbar
- Bei der Herstellung können laterale Dimensionen verkleinert oder vergrößert werden (z.B. durch Diffusion). Daher ist z.B. die wahre Breite $W = W_{\text{drawn}} - W_{\text{offset}}$. W_{offset} ist ein Parameter der Technologie.

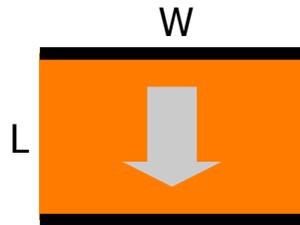
Widerstände mit allgemeiner Form

- Bei allgemeineren Formen muß (im Prinzip) die Poisson-Gleichung gelöst werden.
- **Näherungsweise** kann eine Struktur in einfache Elemente zerlegt werden:

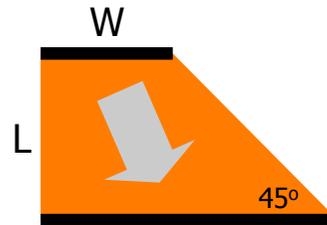


Kompliziertere Geometrien

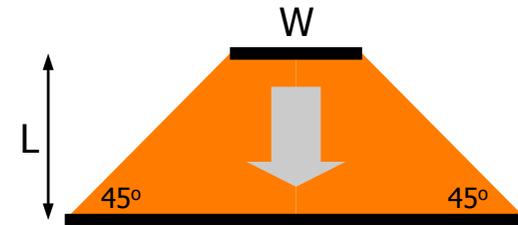
- Für einige einfache Formen gibt es Formeln:



$$R_{\square} \cdot L/W$$

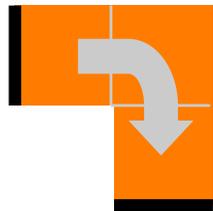


$$R_{\square} \cdot 4L/(L+4W)$$



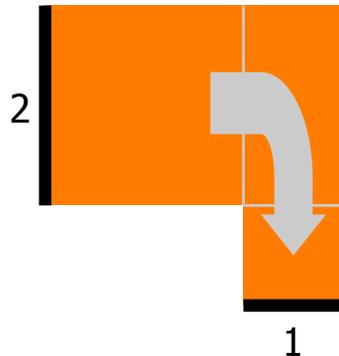
$$R_{\square} \cdot 2L/(L+2W)$$

- Für elementare Ecken etc. gibt es Tabellen mit ungefähren Werten, z.B.:

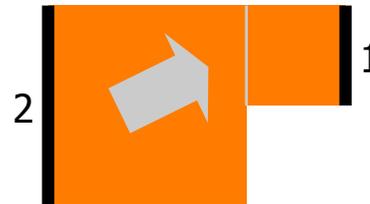


$$\sim 2.5 R_{\square}$$

(2.53...2.65 R_{\square})
(my Tool: 2.52)

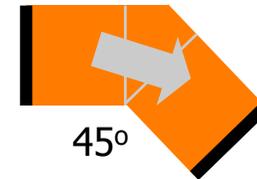


$$\sim 2.55 R_{\square}$$



$$\sim 2.25 R_{\square}$$

(my Tool: 2.29)



$$\sim 2.2 R_{\square}$$

(auch 2.33 R_{\square})



$$\sim 2.96 R_{\square}$$

Große Widerstände

- Um hohe Widerstandswerte zu bekommen, muß man die Widerstände **lang** und **schmal** machen.
- Die Breite kann man durch ein '**dogbone**' (Hundeknochen) - Layout reduzieren.
Achtung: Fehler in W durch W_{offset} wirken sich hier sehr stark aus !

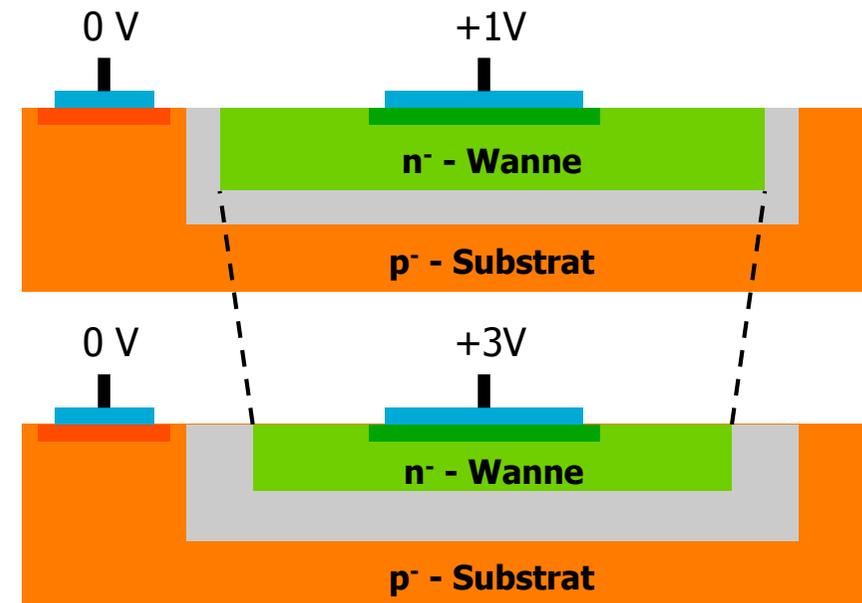


- Lange Widerstände werden gefaltet ('**serpentine**', '**snake**', '**meander**')

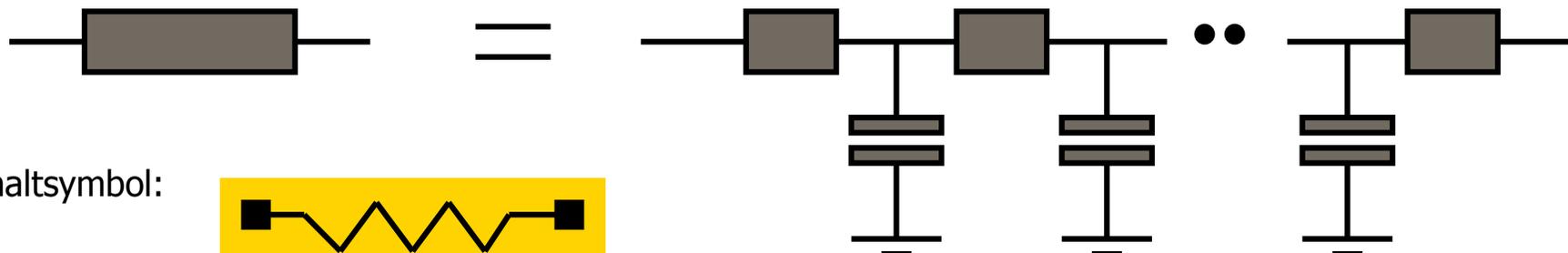


Typische Werte für R_{\square}

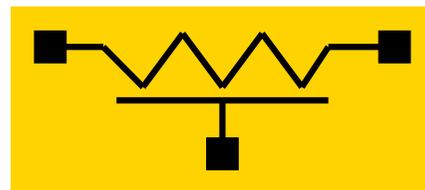
- NWELL 1000 Ω / \square
 - ACTIVE (N+ oder P+) 100 Ω / \square
 - Poly (mit Silicide) 10 Ω / \square
 - Metall 0.1 Ω / \square
- Bei der Auswahl sind außerdem zu beachten:
- **Toleranzen** (Min/Max-Werte der Technologie)
 - **Temperaturabhängigkeit** (z.B. bei NWELL)
 - **Spannungsabhängigkeiten** (Verarmungszonen!)



- Die Widerstände bilden **Kapazitäten!** (NWELL, ACTIVE). Modellierung daher oft als verteiltes RC-Netz:

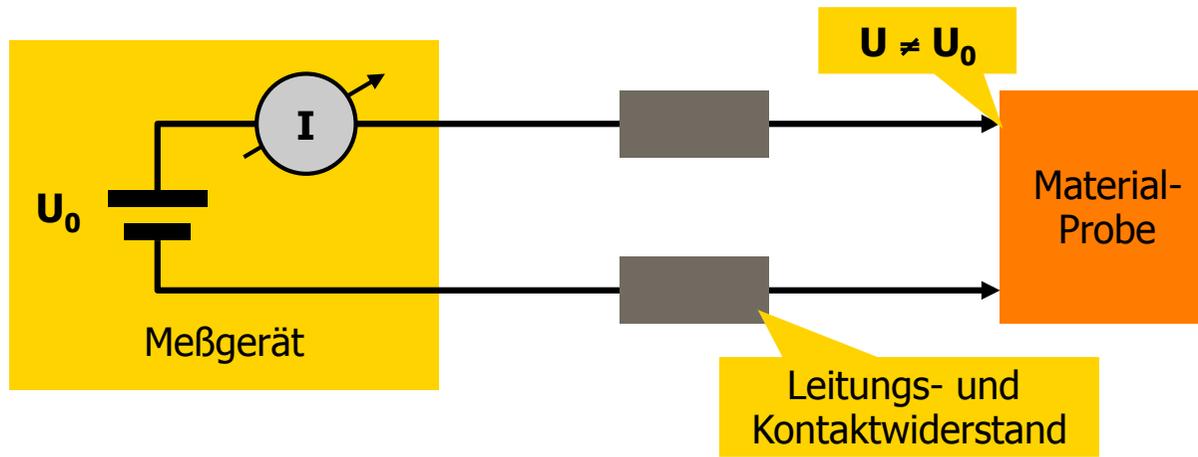


- Schaltsymbol:

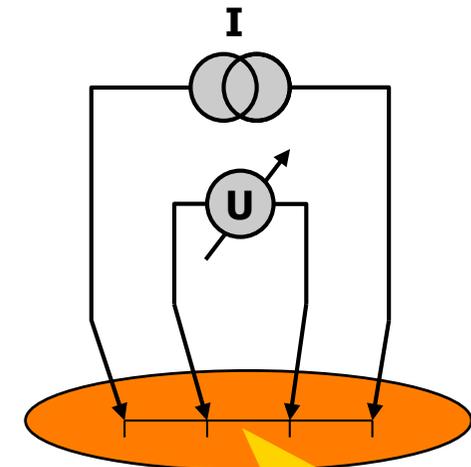
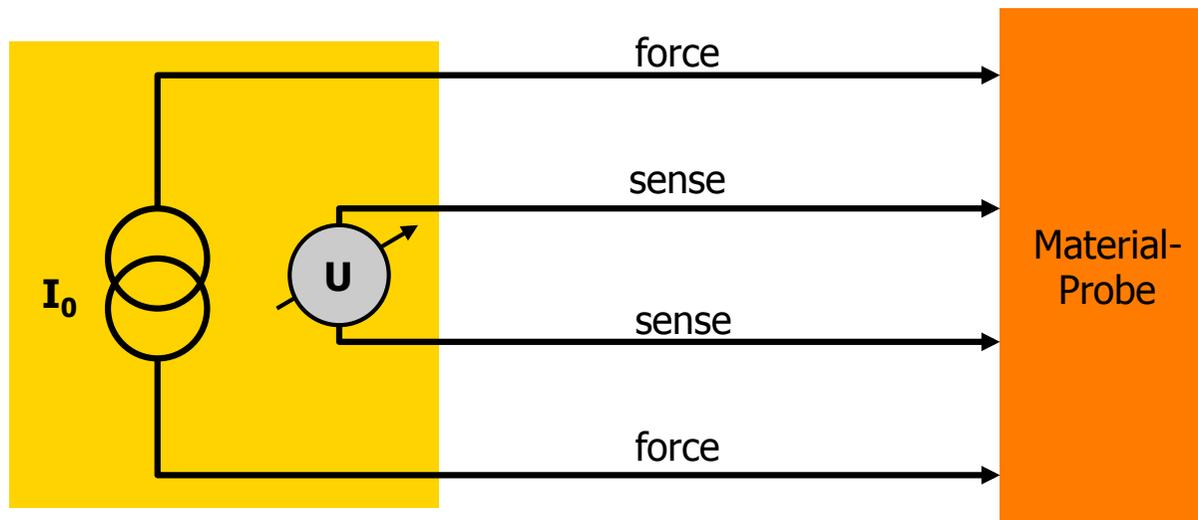


Messung von R

- Problem bei Messung mit zwei Nadeln: Spannungsabfall in den Leitung und Kontaktwiderstand



- klassische Lösung: Vierpunktmessung



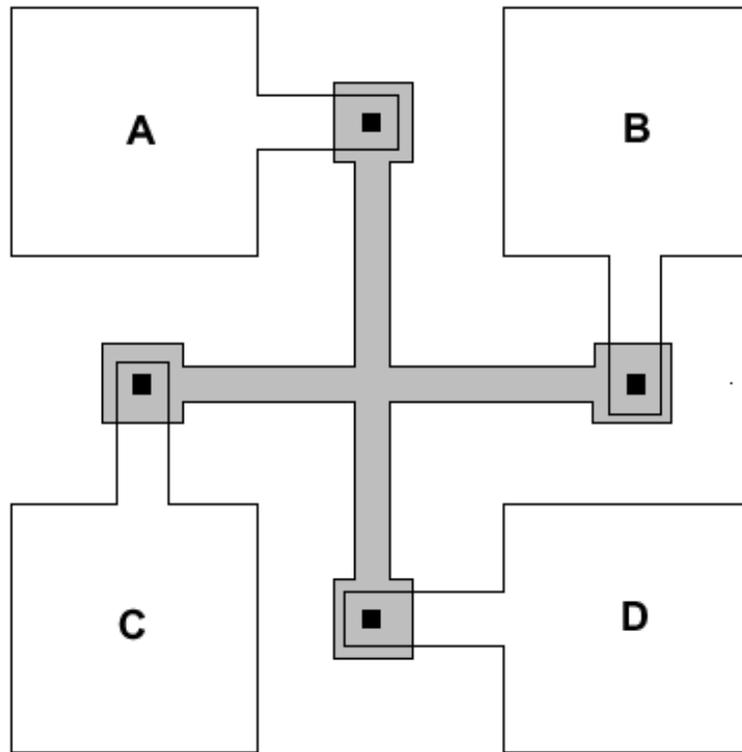
Wird bei 4 äquidistanten Meßspitzen der Widerstand R gemessen, so ist

$$R_{\square} = U/I \cdot \pi / \ln(2)$$

(Die Herleitung erfordert nur elementare Elektrostatik!)

Messung an Teststrukturen

- Mit speziellen Teststrukturen gibt es viele Möglichkeiten
- Man muß aber immer beachten, daß die Geometrie u.U. nicht bekannt ist (Überätzen etc.)
- Beispiel: Das 'Griechische Kreuz':



A.J.Walton: Microelectronic Test Structures

- Strom wird über AB aufgeprägt
- Spannung wird an CD gemessen
- $R_{\square} = U_{CD} / I_{AB} \times \pi / \ln(2)$
- Verbesserung: Mehrere Messungen mit vertauschten Anschlüssen und Mittelung
- Die 'Arme' müssen deutlich länger als das mittlere Quadrat sein
- Fehler dann bis 0.1%
- Der Meßstrom darf nicht zu hoch sein, damit sich die Struktur nicht zu stark erwärmt

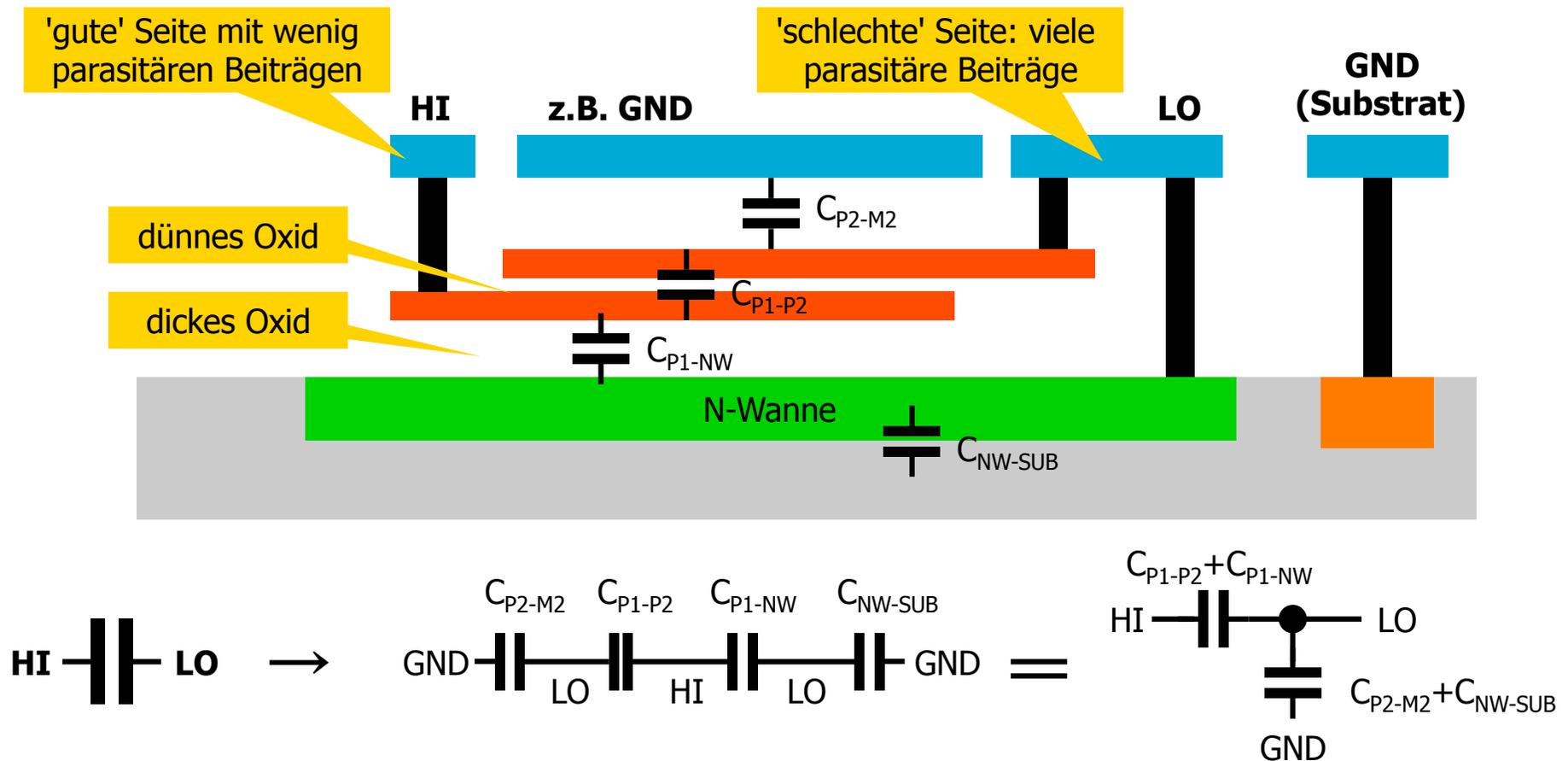
Kondensatoren

Kondensatoren

- Jedes Lagenpaar bildet Kapazitäten
- Man hat wieder Anteile von **Fläche (Area)** und **Rand (Periphery)**
- **Hohe** Kapazitäten hat man durch das Gate-Oxid $C \sim 5 \text{ fF}/\mu\text{m}^2$ (in $0.35\mu\text{m}$)
Diese Kapazität ist **spannungsabhängig** (s. VL Transistor: MOS Struktur)!
Der Transistor muß 'on' sein ($V_{GS} > V_T$) oder in Akkumulation
- **Lineare** Kapazitäten bekommt man immer parasitär als Poly-M1, M1-M2, ...
 - nur kleine Werte: **Area** $\sim 0.03 \text{ fF}/\mu\text{m}^2$, **Peripherie (engl. ‚FRINGE‘)** $\sim 0.04 \text{ fF}/\mu\text{m}$
 - z.B. $1 \mu\text{m}$ breite Leiterbahn: $(0.03 + 2 \times 0.04) \text{ fF}/\mu\text{m}$
- Manche Technologien bieten einen dünnen Zwischenisolator für große, lineare Kapazitäten (z.B. **Poly1 – Poly2**, oder **MIM**-Cap 'Metal-Insulator-Metal' = **MMC** (Metal.Metal.Cap))
 - Dann $\sim 1 \text{ fF}/\mu\text{m}^2$ oder mehr
- Jede Lage trägt zu (unerwünschten) **parasitären Kondensatoren** bei!

Poly1-Poly2 Cap

- Hier muß man beachten, daß es keine Kontakte von Poly1 nach Poly2 gibt (Herstellungstechnologie!)
- Zur Abschirmung zeichnet man oft eine N-WANNE unter die Kapazität.
- Manchmal deckt man die Kapazität mit M1 ab, dann ist klar definiert, wohin die parasitären Beiträge gehen.
- Die parasitären Kapazitäten sind dann sehr unterschiedlich. Man muß den Kondensator 'richtig herum' in die Schaltung einbauen!

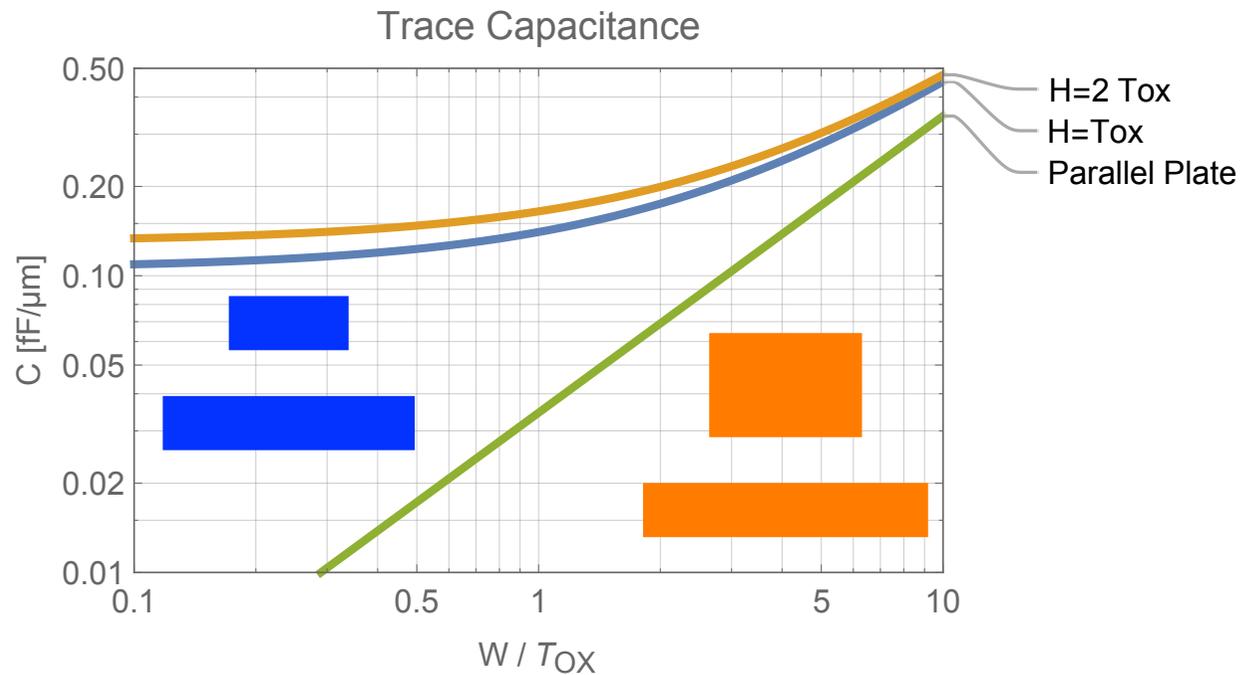
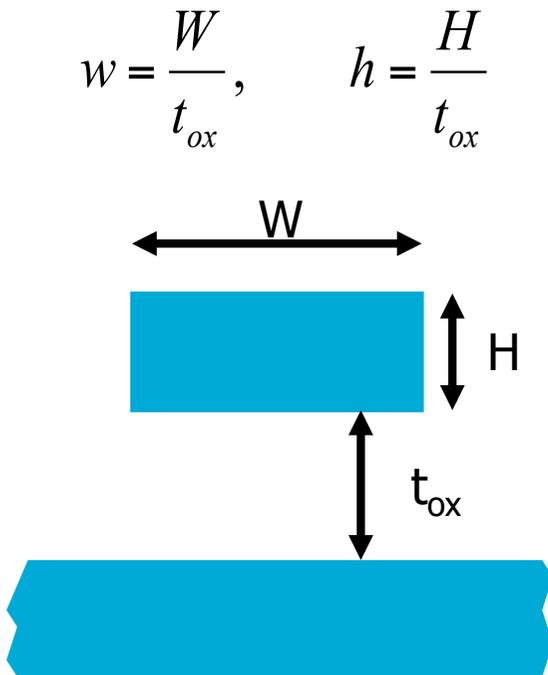


Kapazität von Leiterbahnen

Für $W/H > 0.5$ gilt ($\sim 10\%$ Genauigkeit):

$$\frac{C}{l} \approx \varepsilon \varepsilon_0 \left\{ w - \frac{h}{2} + \frac{2\pi}{\ln\left(1 + \frac{2}{h} + \frac{2}{h} \sqrt{1+h}\right)} \right\}$$

Yuan, Trick, 1982



Kapazität von Leiterbahnen

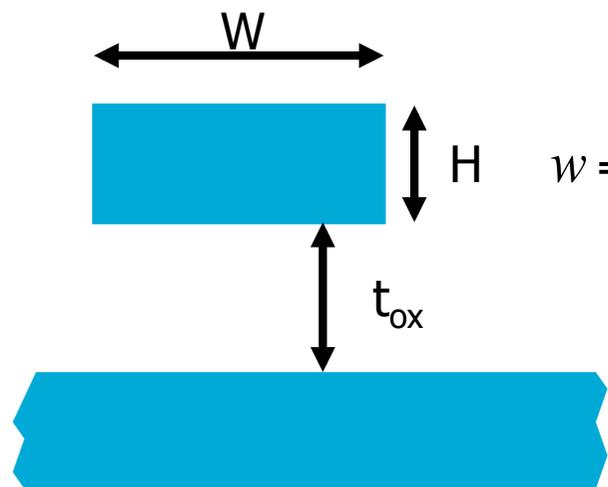
- Andere (empirische) Näherungsformeln:

$$\frac{C}{l} \approx \varepsilon \varepsilon_0 \left\{ 1.15 \cdot w + 2.80 \cdot h^{0.222} \right\}$$

Sakurai, Tamaru, 1983

$$\frac{C}{l} \approx \varepsilon \varepsilon_0 \left\{ w + 0.77 + 1.06 \cdot w^{0.25} + 1.06 \cdot h^{0.5} \right\}$$

Meijs, Kokkema, 1984



$$w = \frac{W}{t_{ox}}, \quad h = \frac{H}{t_{ox}}$$

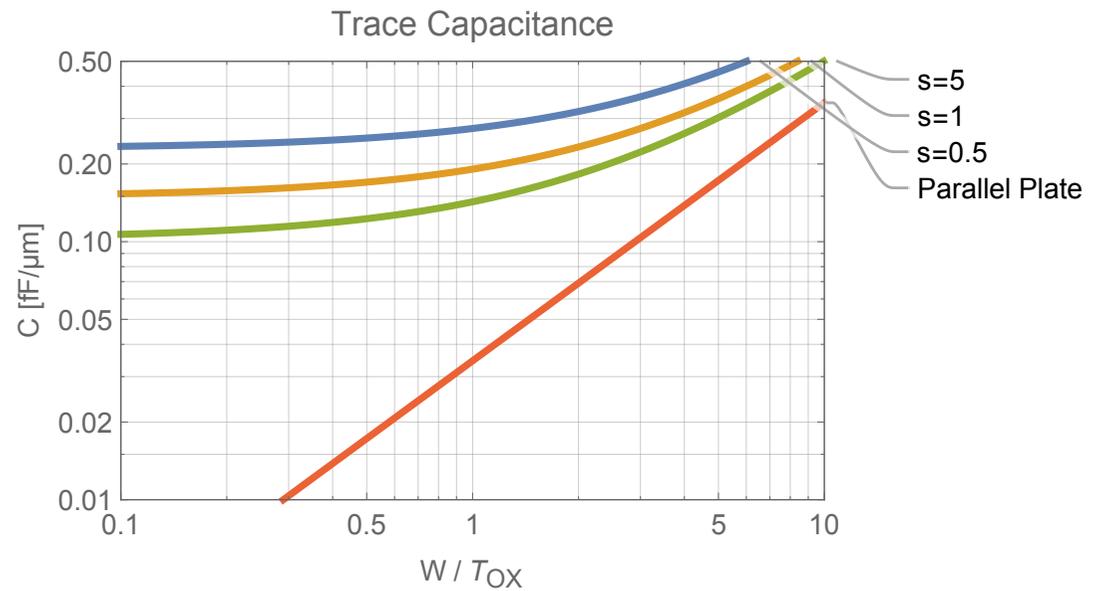
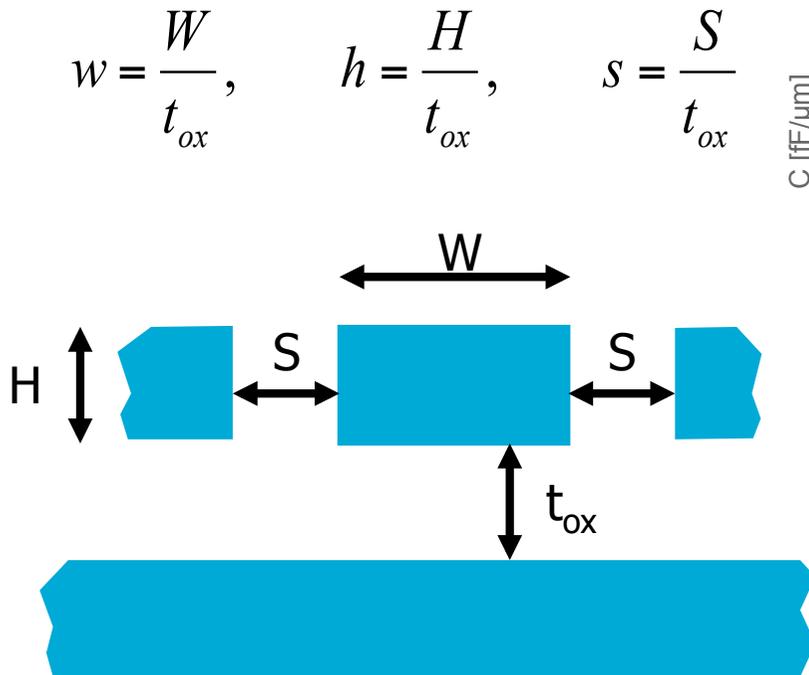
Sehr genau,
einfache
Exponenten

$$\varepsilon_{\text{SiO}_2} \varepsilon_0 \approx 3.9 \cdot 8.85 \times 10^{-12} \text{ F/m} = 34.52 \text{ aF}/\mu\text{m}$$

Kapazität von Leiterbahnen

Mit Nachbarn:

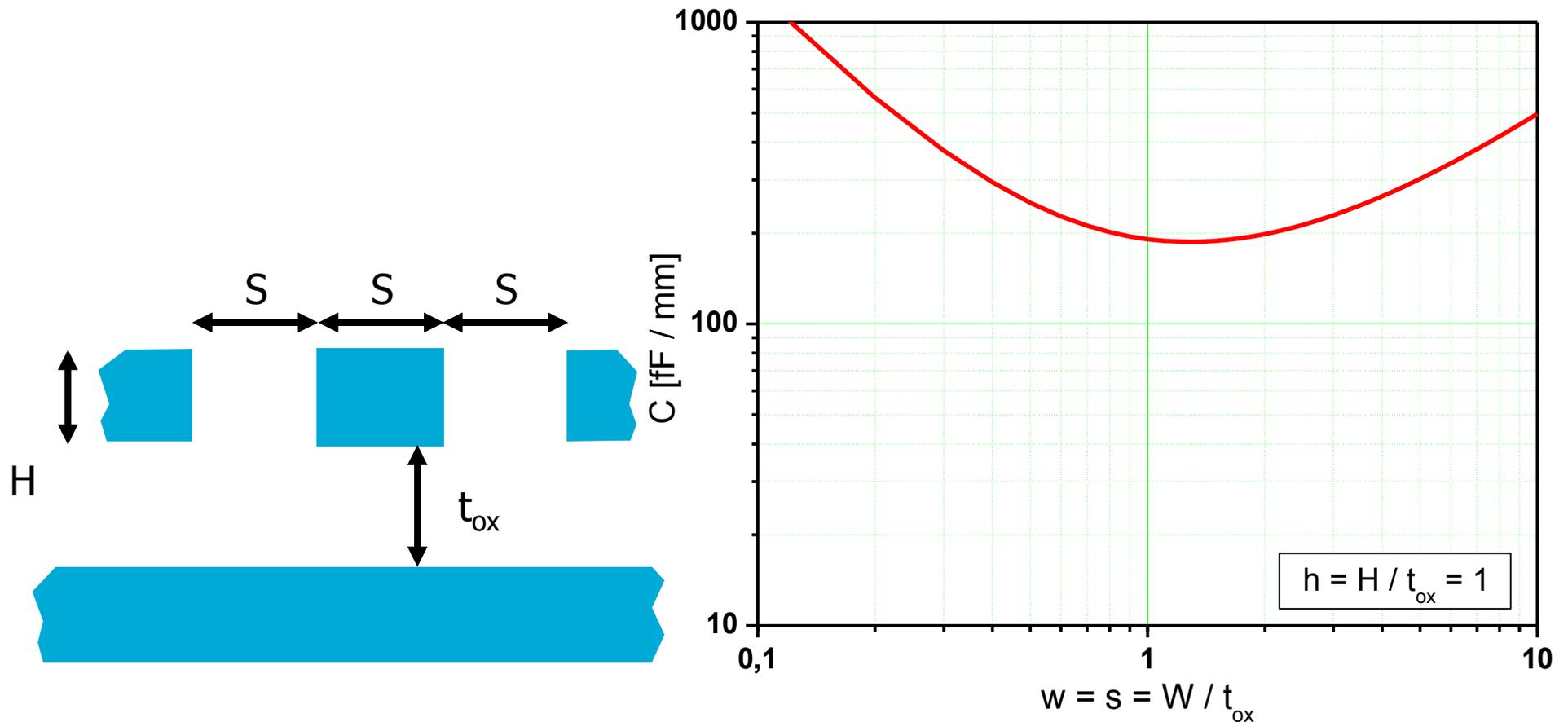
$$\frac{C}{l} \approx \epsilon \epsilon_0 \left\{ 1.15w + 2.8h^{0.222} + \frac{0.06w + 1.66h - 0.14h^{0.222}}{s^{1.34}} \right\}$$



Kapazität von Leiterbahnen

Bei der Konfiguration Breite = Abstand gibt es ein minimales C!

Spezialfall $W = S$

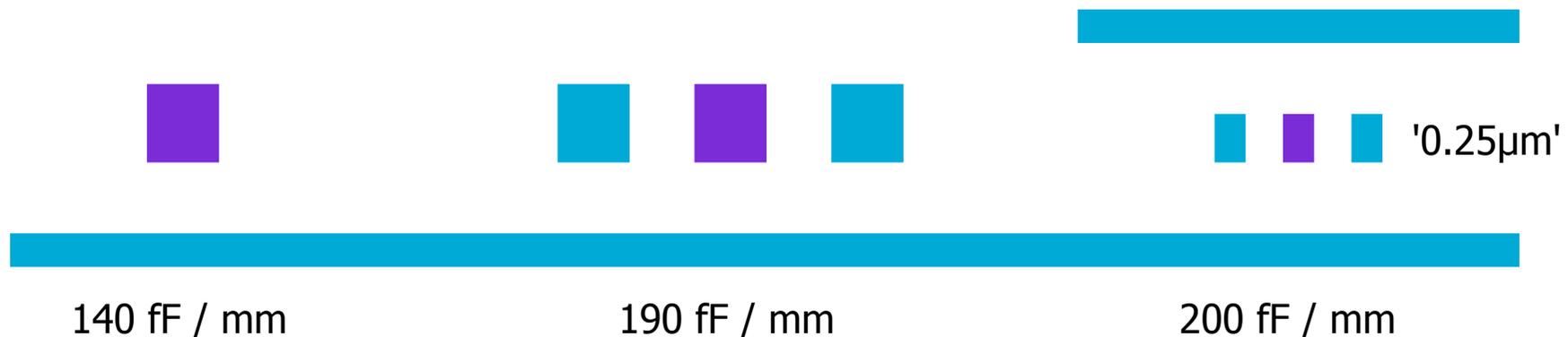


Streifelder - Zusammenfassung

- Plattenkondensator und Streiffelder tragen zur Leitungskapazität bei
- **Bei schmalen Leitungen dominieren die Streifelder**
- Bei 0.25 μm Technologie und darunter dominieren oft die Cs zu den Nachbarn ('dicke Metallagen')
- Nachbarleitungen und darüberliegende Metallebenen erhöhen die Kapazität
- Hier lohnt es sich daher, den Abstand zu erhöhen!
- Typische Werte (0.8 μm Technologie): $C_{\text{area}} = 35\text{aF}/\mu\text{m}^2$, $C_{\text{finge}} = 51\text{aF}/\mu\text{m}$
- Wenn C genau definiert sein muß, sollten die Streifelder genau definiert sein:

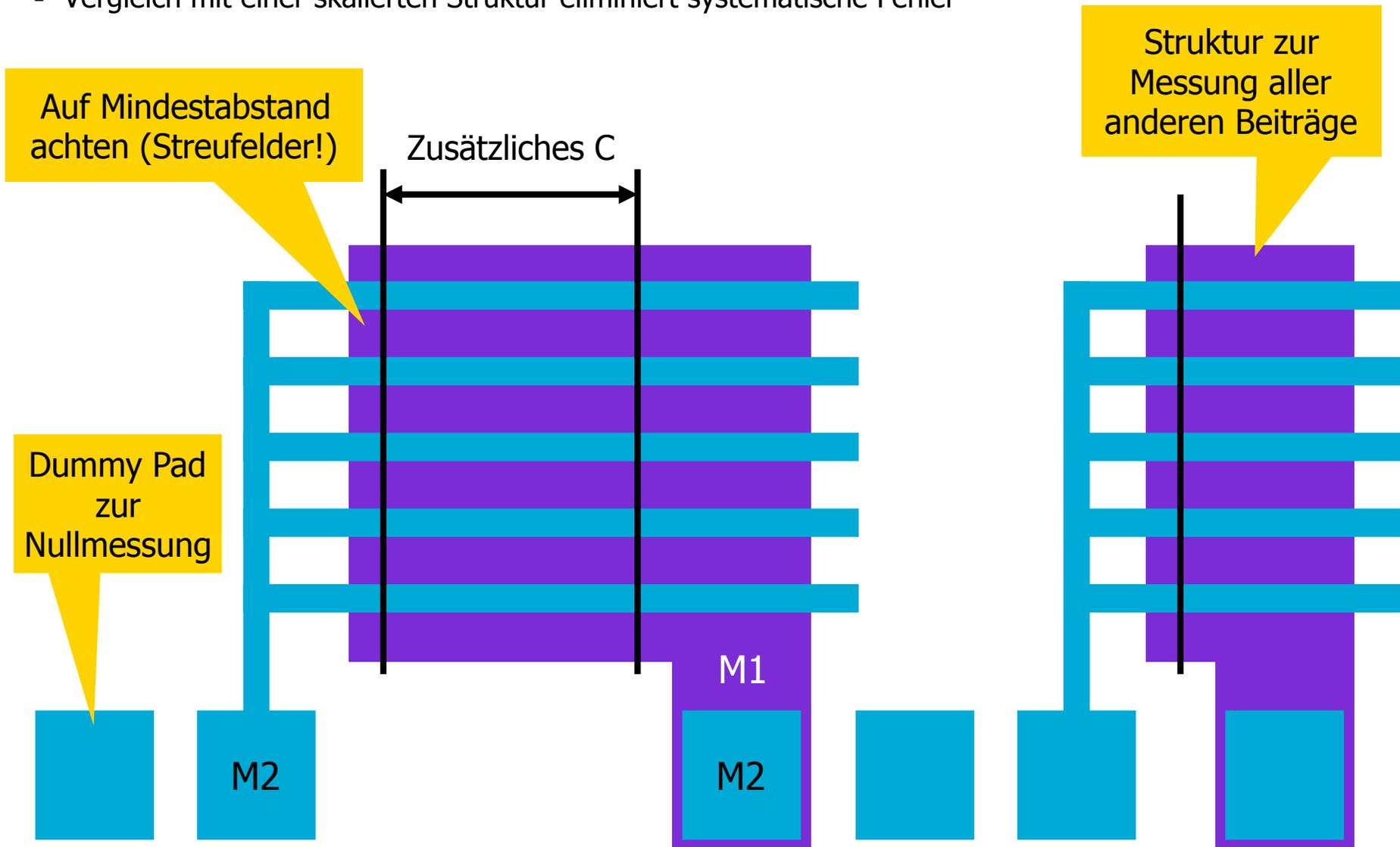


- Typische Topologien:



Layout zur Kapazitätsmessung am Spitzenmeßplatz

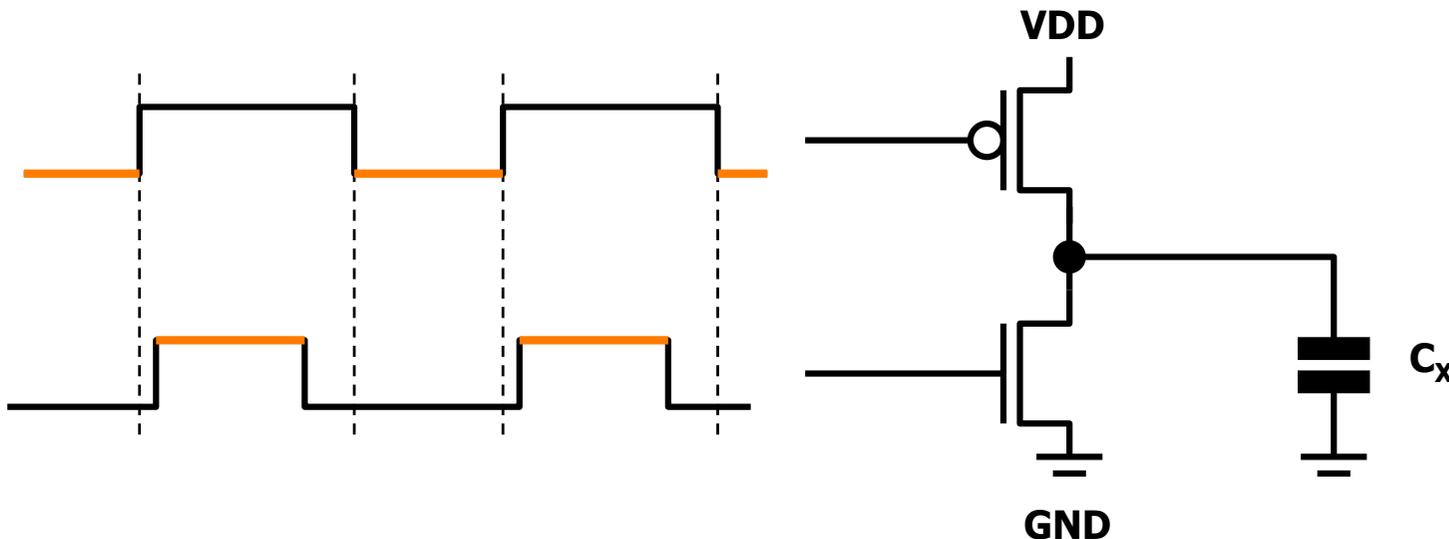
- Parallelschaltung vieler Elemente reduziert den Meßfehler
- Vergleich mit einer skalierten Struktur eliminiert systematische Fehler



Kapazitätsmessung mit Ladungspumpe

- Sehr einfache und genaue Methode:
- Die unbekannte Kapazität wird periodisch auf Spannung aufgeladen und entladen (nicht überlappende Takte PHI1 und PHI2)
- Der mittlere Strom wird gemessen.
Pro Zyklus wird die Ladung $VDD \times C_x$ von VDD nach GND transportiert.

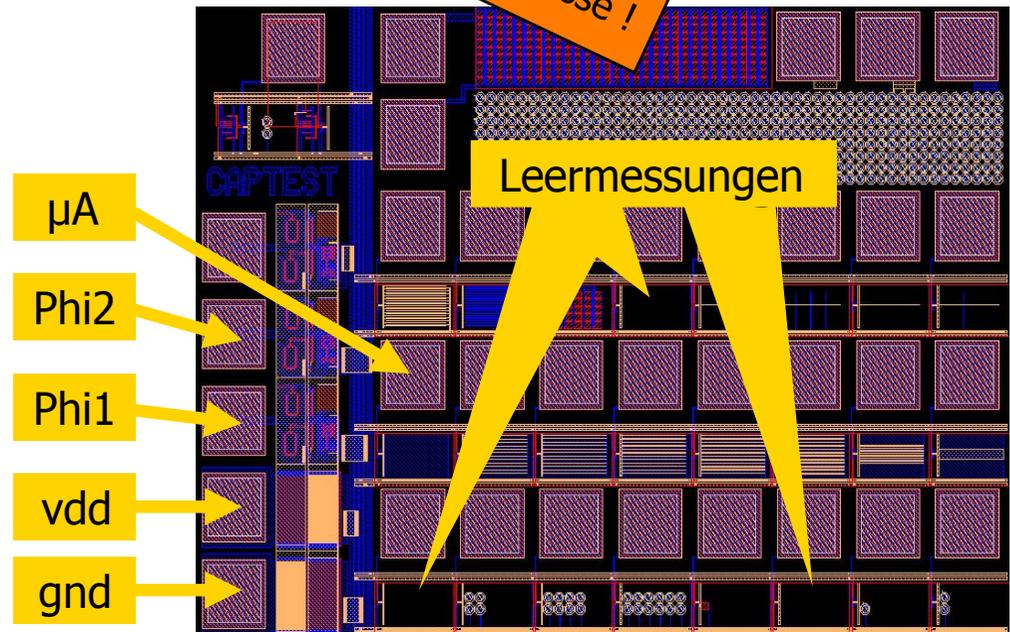
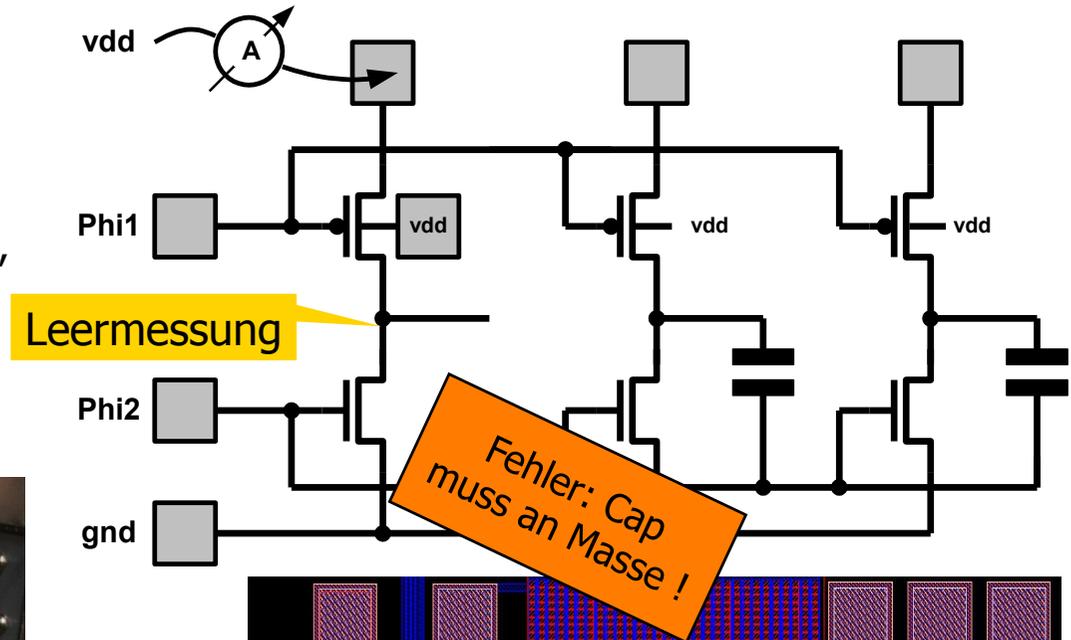
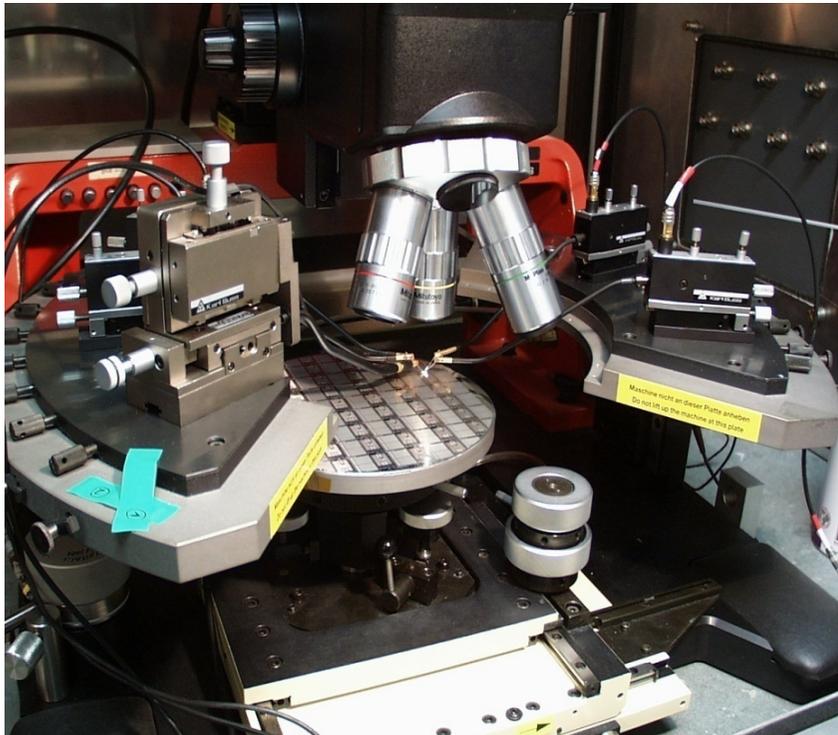
Der mittlere Strom ist also $I = f \times VDD \times C_x$



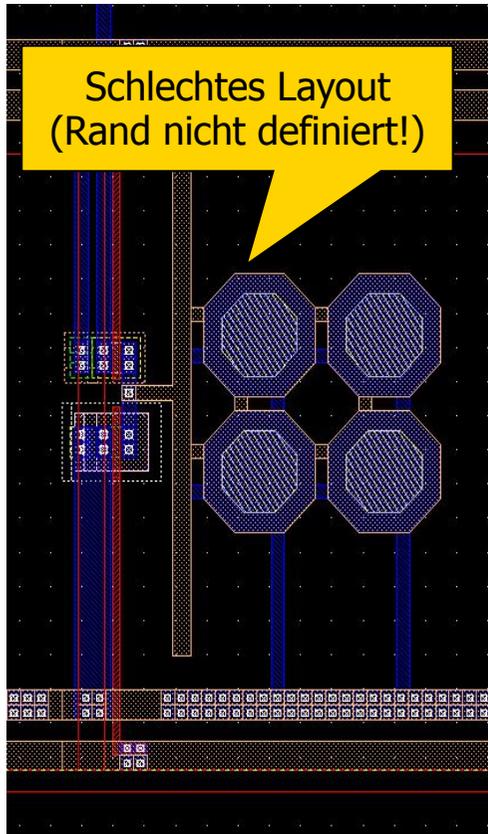
Referenz: J.C.Chen et al., An On-Chip Interconnect Capacitance Characterization Method with Sub-Femto-Farad Resolution, IEEE Trans. on Semiconductor Manufacturing, Vol. 11, No.2, May 1998.

Messung auf Wafer am Spitzenmeßplatz

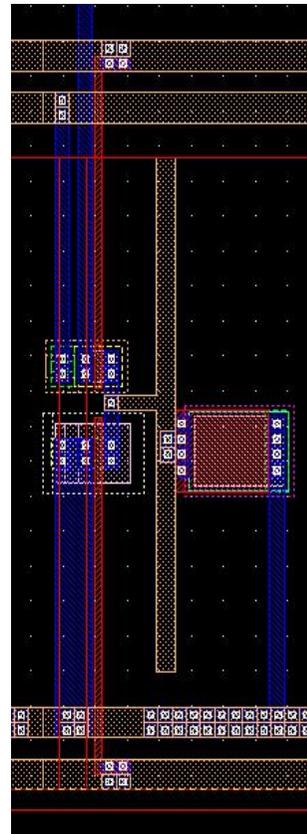
- Messung mit 5 Nadeln
- Nur **eine** Nadel wird bewegt
- Geräte:
Pulsgenerator, Amperemeter
- Alle parasitären Elemente (Drain-Kapazitäten, Leitungen) werden durch Leermessungen ermittelt und abgezogen.



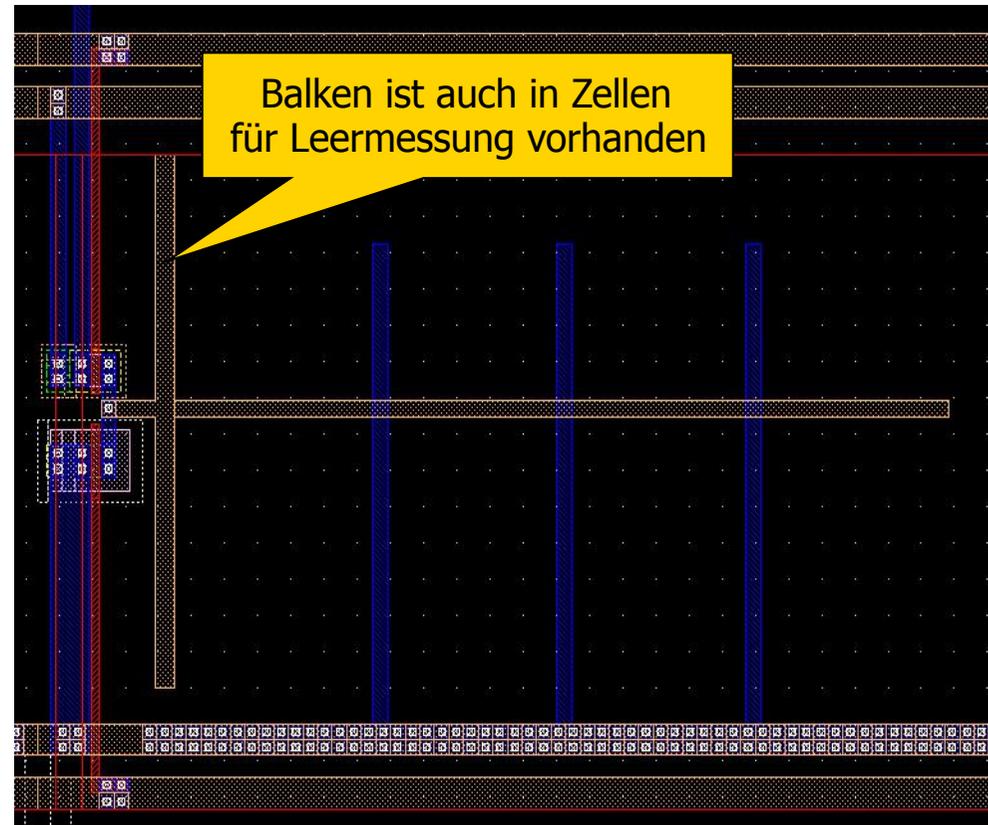
Layoutbeispiele



M1-M2 Kapazitäten



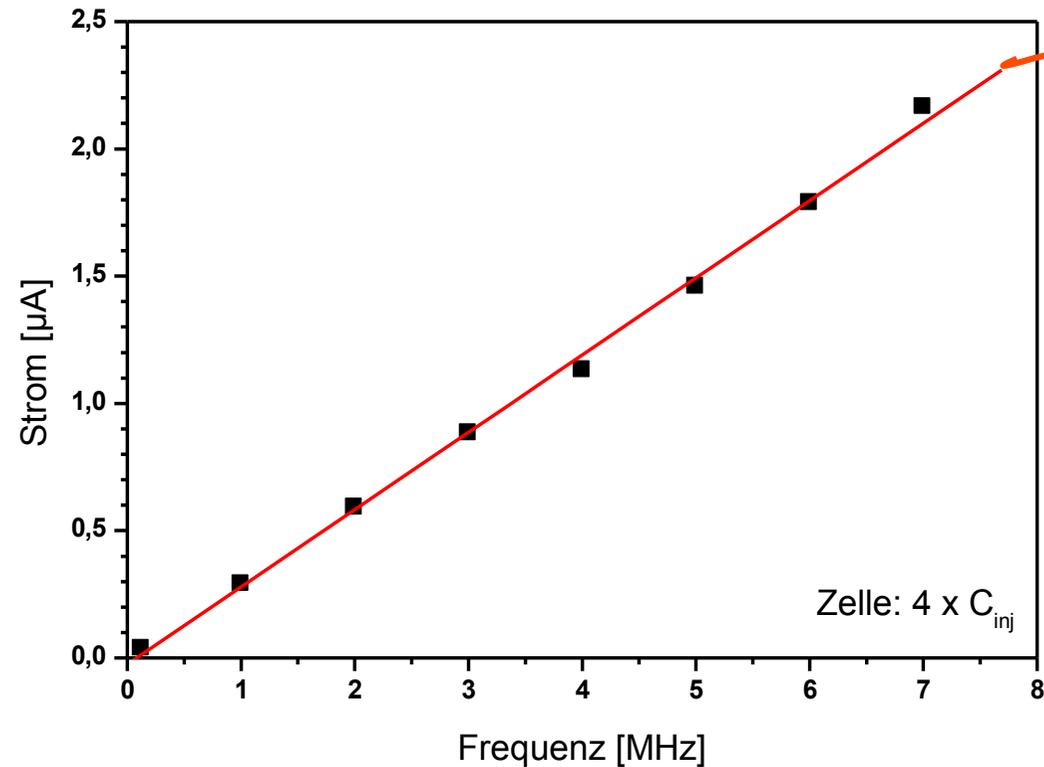
Poly1-Poly2
Kapazität



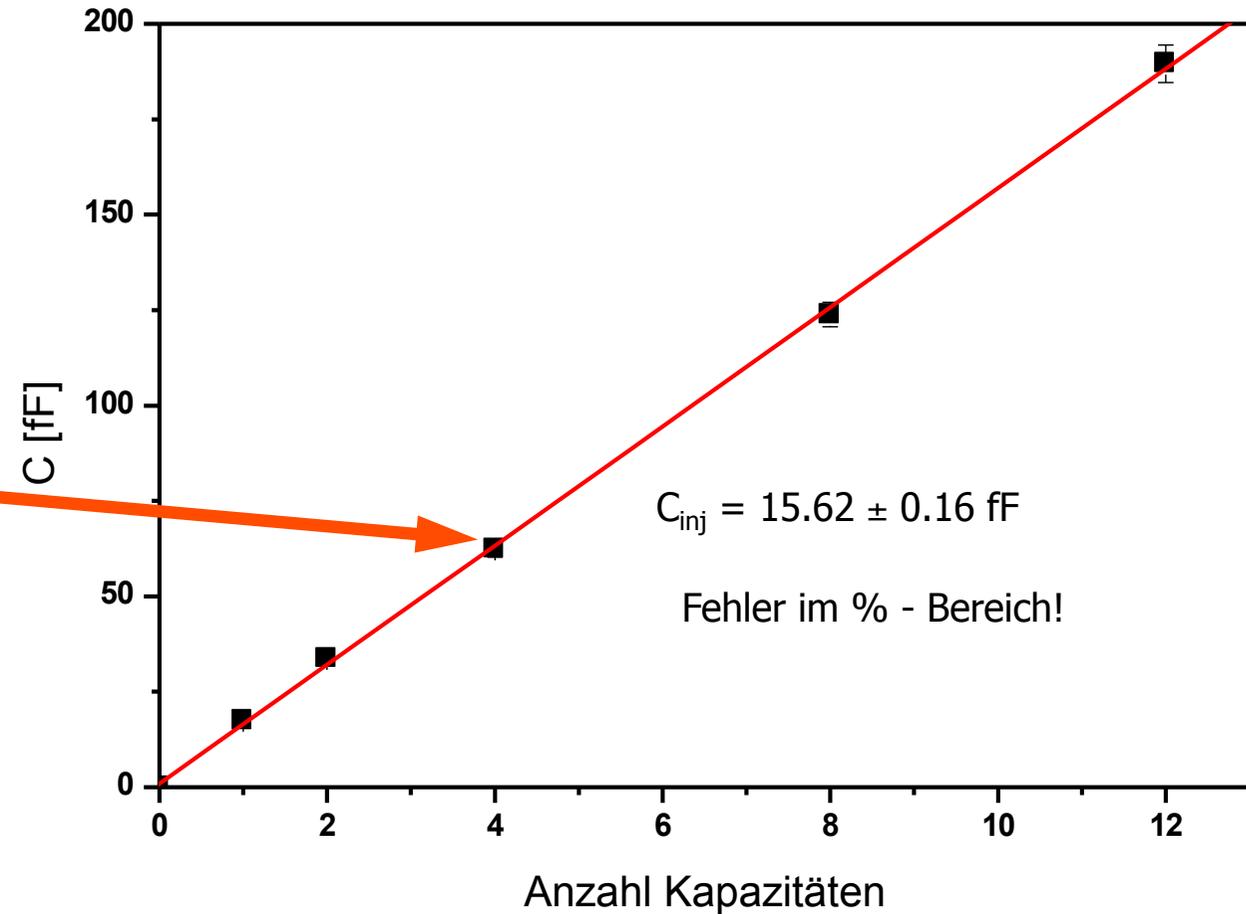
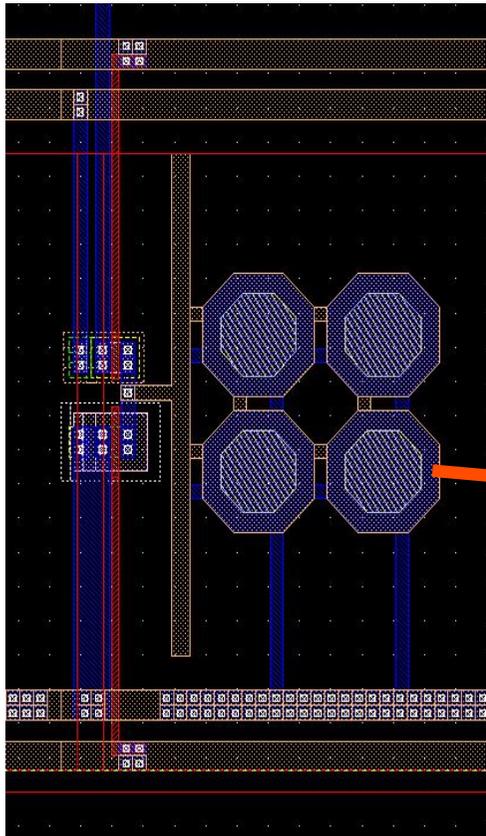
Drei M1 - M2 Kreuzungen
(auch Zellen mit 1 oder 2 Kreuzungen)

Messung: Strom = f(Frequenz)

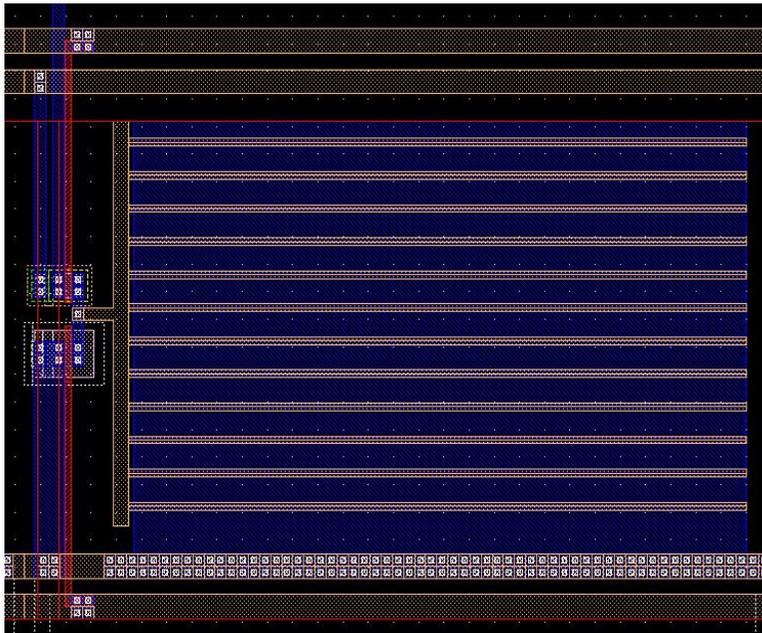
- Messung bei verschiedenen Frequenzen reduziert den Fehler: $I = v_{dd} \cdot C \cdot f \Rightarrow C = \frac{1}{v_{dd}} \cdot \frac{I}{f}$
- Zur Kontrolle: Messung bei verschiedenen Versorgungsspannungen v_{dd}
- Die Messung aus leeren Zellen wird abgezogen (hier $38.9 \pm 0,7$ fF).
- Mit mehreren leeren Zellen kann der statistische Fehler abgeschätzt werden (Streuung der parasitären Kapazitäten)



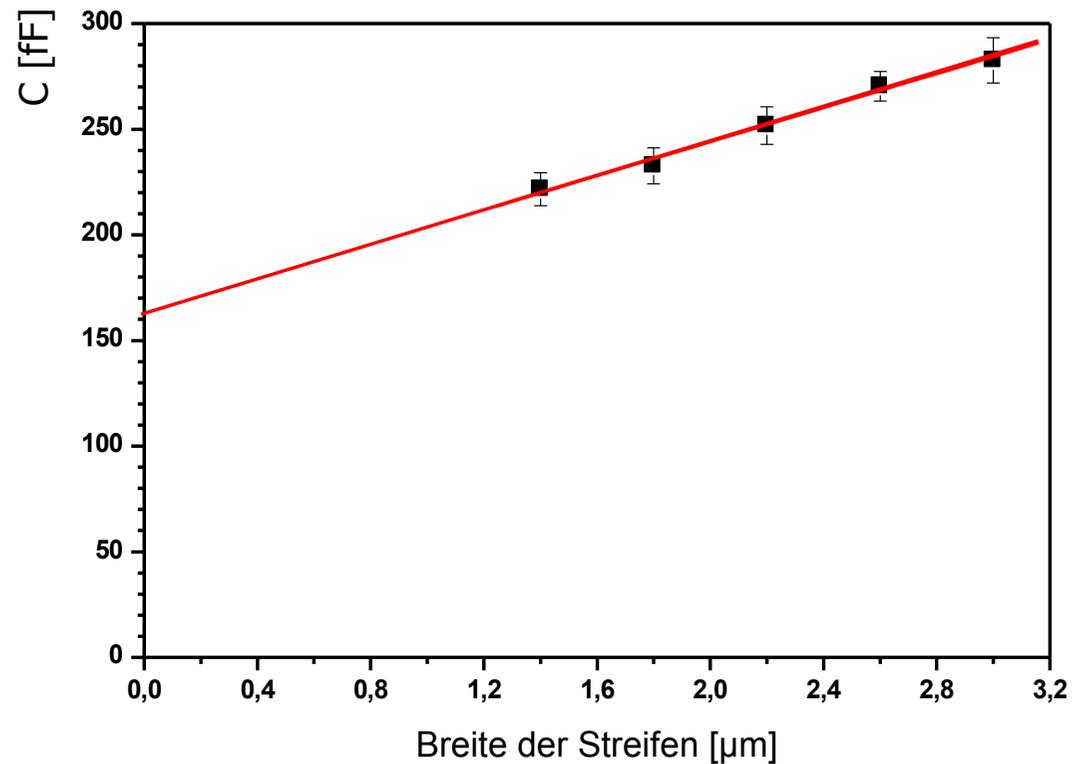
Beispiel für Ergebnisse: Injektionskapazitäten



Messung Parallelplatten-C und Streufeld-C



12 Streifen mit unterschiedlicher Breite
(121.6 μm lang, 6.0 μm Raster)



Steigung $\Rightarrow C_{\text{area}}$ Rechnung: = **28.3 aF / μm² ± 8%**
28.8 aF / μm²

($\epsilon=3.9$, $t_{\text{ox}}=1.2\mu\text{m}$)

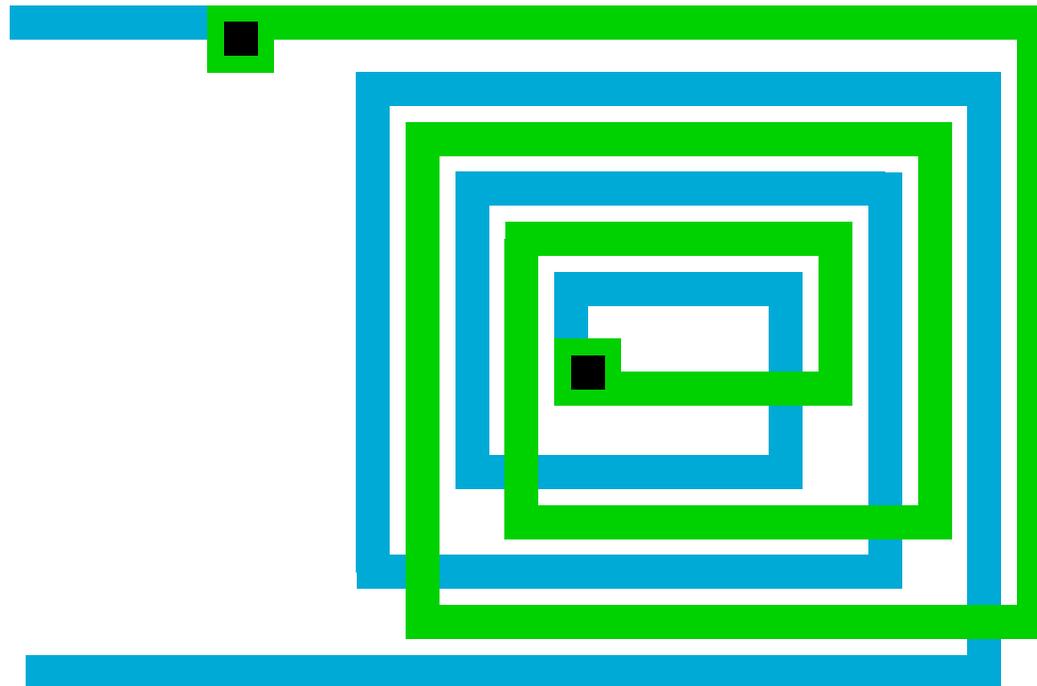
Achsenabschnitt $\Rightarrow C_{\text{fringe}}$ Rechnung: = **55.3 aF / μm ± 4%**
50.8 aF / μm

($\epsilon=3.9$, $t_{\text{ox}}=1.2\mu\text{m}$, $H=1.05\mu\text{m}$)

Spulen

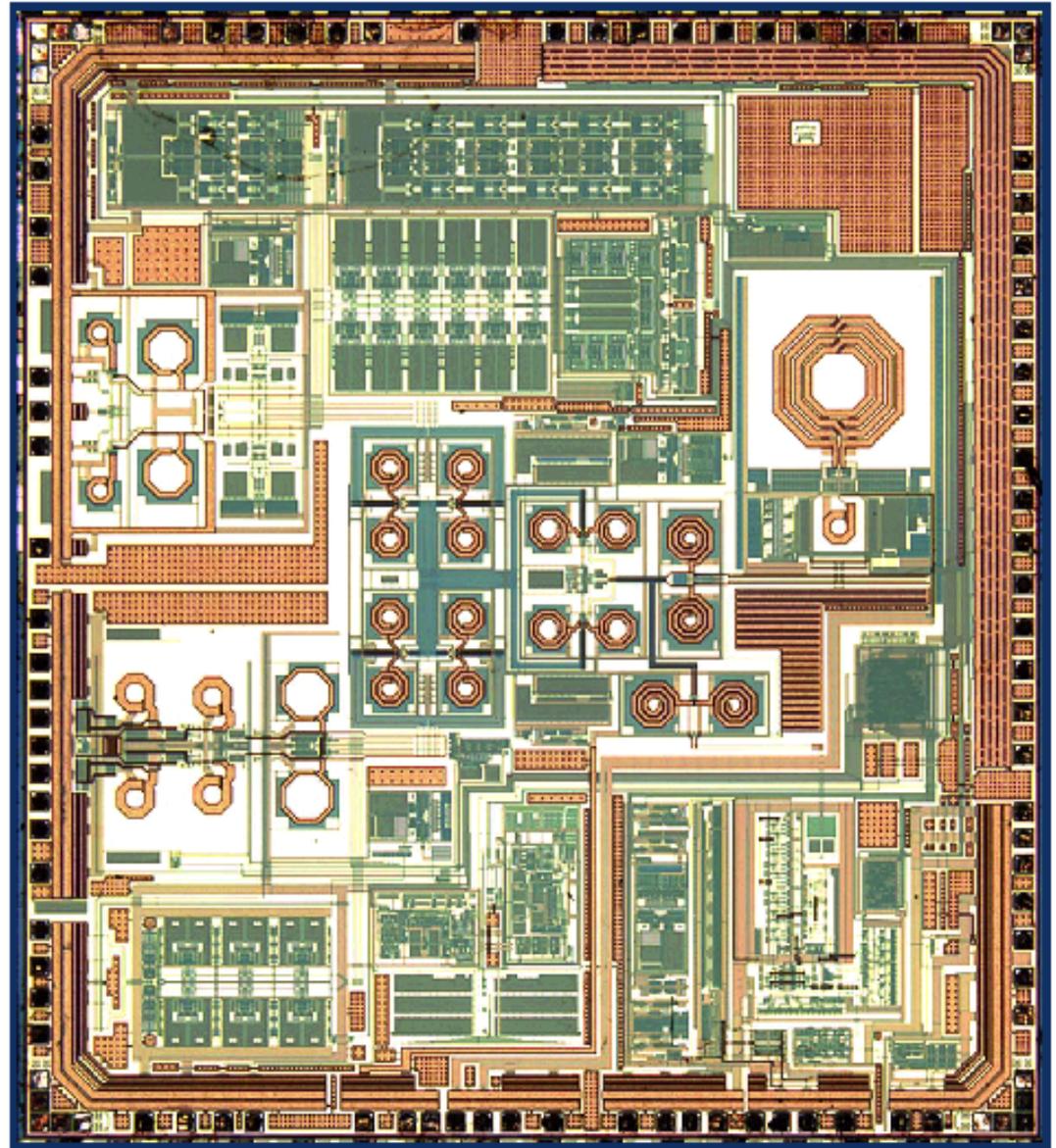
Spulen

- Sind möglich, seit viele Metalllagen verfügbar sind
- Problem: Qualität (Güte der Spule) wird durch parasitäre Kapazitäten verschlechtert
- Oft Wicklungen in mehreren Lagen
- Bevorzugt: Obere Lagen, da diese oft dicker sind und daher der Widerstand kleiner ist. Außerdem sind parasitäre Kapazitäten kleiner.
- Beispiel mit 2 Lagen:



Spulen

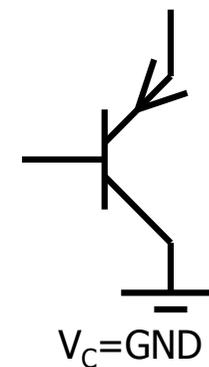
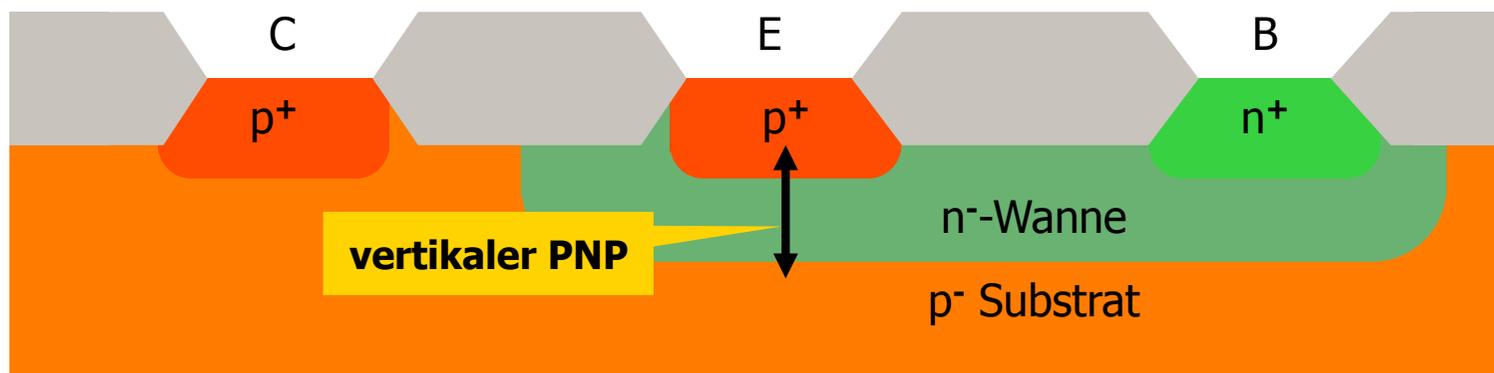
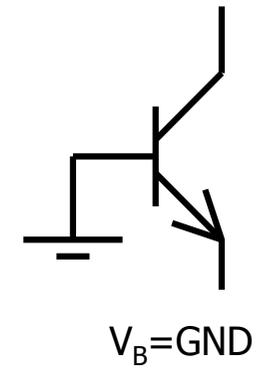
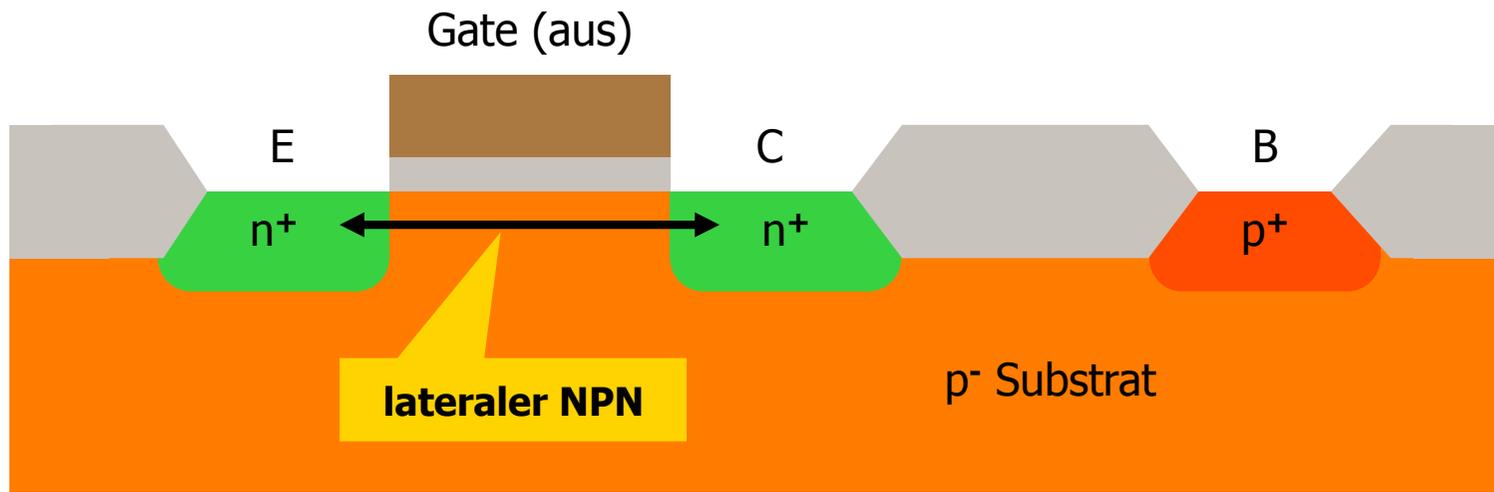
- Werden im HF Bereich (einige GHz) benötigt:
 - Telecom
 - WLAN, (Bluetooth, Zigbee)...
- Baugruppen:
 - LNAs (Low noise Amplifier)
 - VCO (Voltage Controlled Oscillator)
 - PLL (Phase Locked Loop)



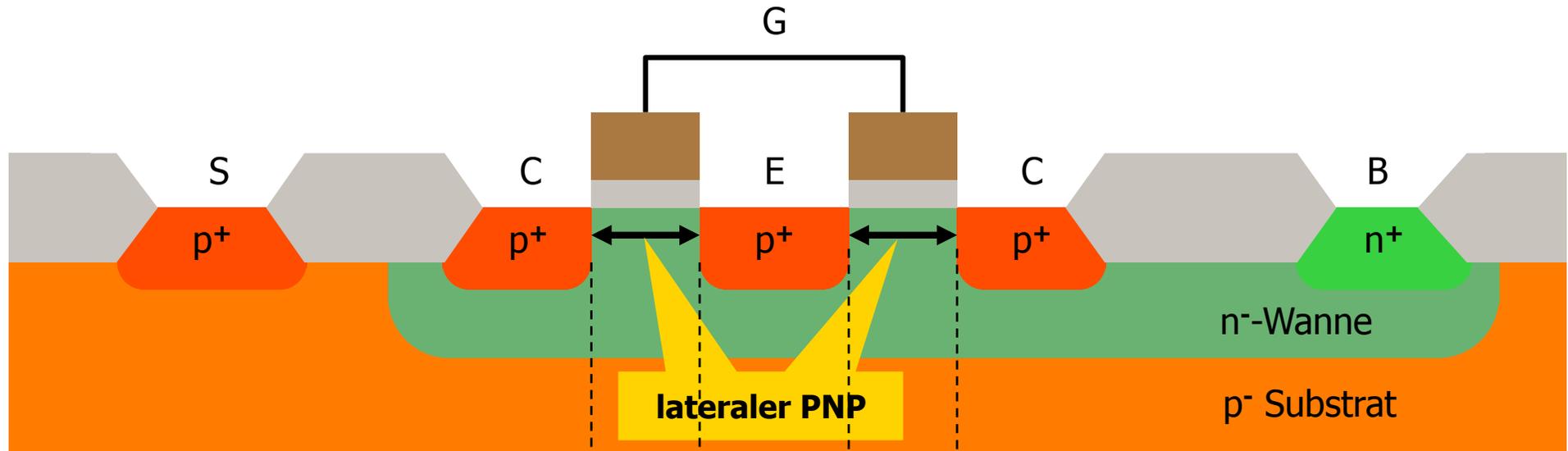
Broadcom BCM2050 WLAN Chip
Quelle Photo: TechInsight

Bipolare Transistoren (in CMOS)

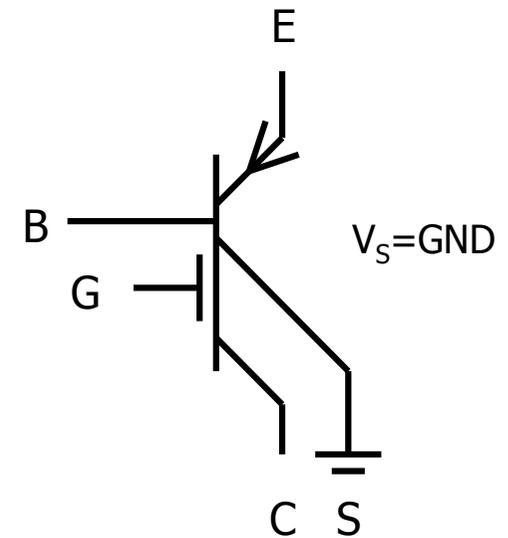
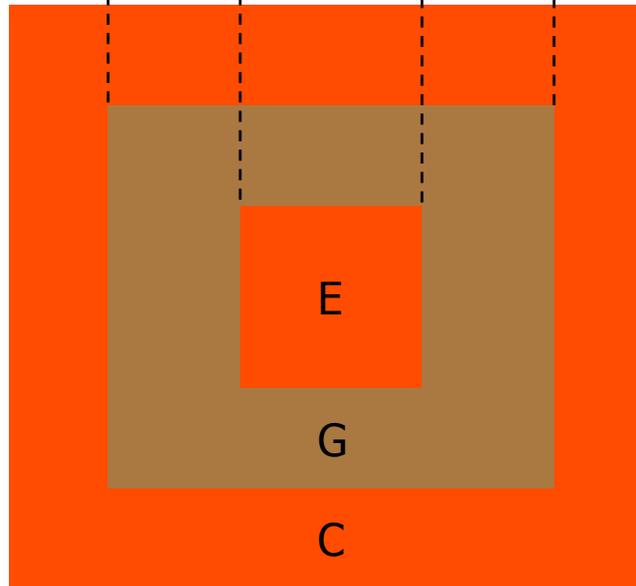
Parasitäre Bipolartransistoren in CMOS



Lateraler pnp Bipolartransistor



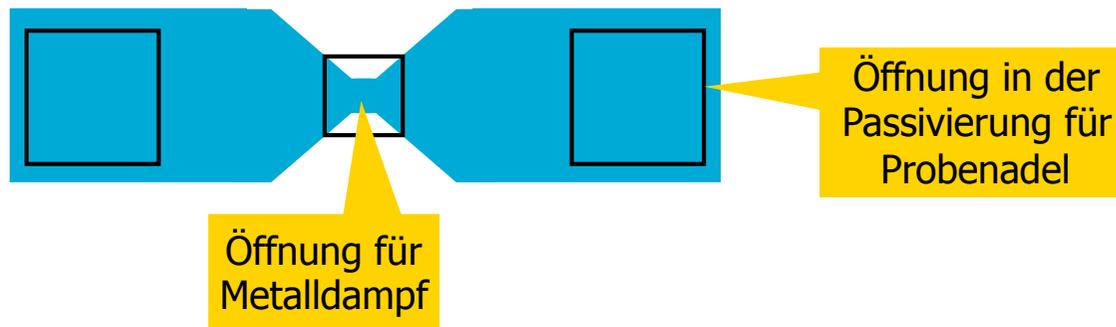
- Analog zum lateralen npn, aber in einer N-Wanne
- Dadurch weniger Einschränkungen in den Spannungen (Basispotential kann frei gewählt werden)
- Es ist immer auch ein vertikaler pnp beteiligt
- Schaltsymbol daher kompliziert. (Gate hier explizit eingezeichnet, es wird meist ausgeschaltet)



Fuses (Sicherungen)

Fuses ('Sicherungen') zum Trimmen

- Zum einmaligen, festen Programmieren von Chips (wenige Bits) oder zum Feinjustieren (Trimmen) werden manchmal Sicherungen benutzt, die beim Programmieren durchgebrannt werden können
- **Metal-fuses:** ein kurzes Stück Leiterbahn, das mit einem hohen Strom verdampft werden kann
 - Widerstand muß klein genug sein, damit der Strom bei wenigen Volt 'Brenn'-Spannung zur Verdampfung des Metalls ausreicht (benötigt werden einige 100mA für einige Millisekunden)
 - Eine Öffnung in der Passivierung (oberste Lage) läßt das verdampfende Metall entweichen, sonst besteht die Gefahr, daß eine unterbrochene Verbindung mit der Zeit wieder 'zuwächst'



- Manchmal auch **Poly-Fuses:** Weniger geeignet, da Widerstand höher und Schmelzen des Si schwieriger
- **Zener-Zapping:** Zener-Dioden können so überlastet werden, daß sich ein **Kurzschluß** bildet
 - Zerstörung des Kristallgitters!
 - Keine Öffnung in der Passivierung nötig, daher auch keine Kontaminierung des ICs möglich
 - Meist nur in speziellen Bipolartechnologien verfügbar

Matching

Matching

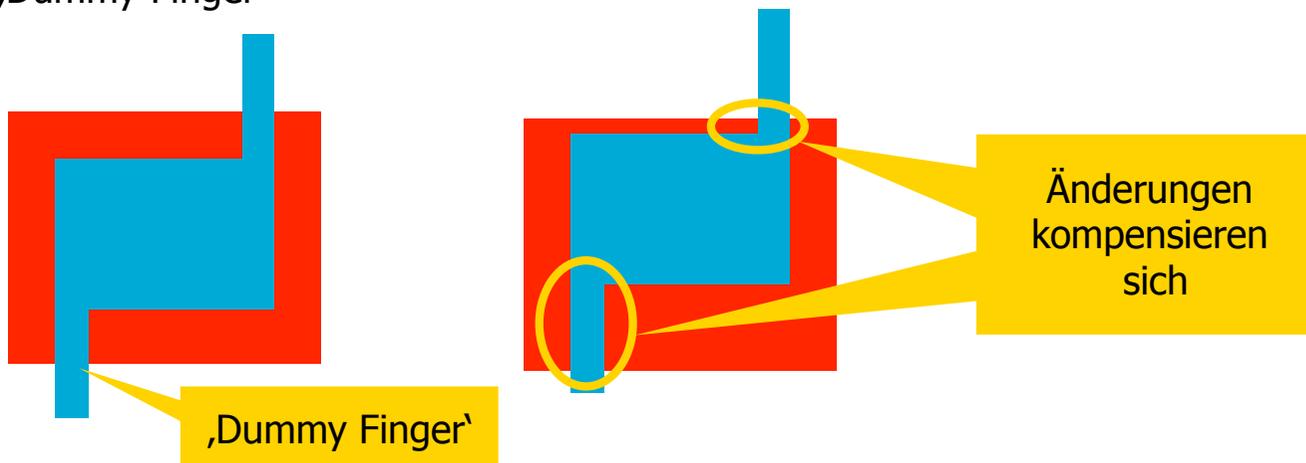
- Zwei identisch gezeichnete Bauteile (z.B. FETs) verhalten sich nicht identisch:
 - zufällige Variationen in der x- und y-Dimension und in LD, WD
 - zufällige Variationen in der Oxiddicke, der Dotierung, ...
 - unterschiedliches Ätzverhalten durch unterschiedliche Umgebung
 - unterschiedliche Temperaturen (Einfluß auf Halbleiterbauelemente, Thermoelektrischer Effekt: ca. 0.4mV/K)
- Sollen sich unterschiedliche Bauteile identisch verhalten (gleicher Widerstand, Kapazität, Strom, Schwellenspannung,...) so muß man:
 - identische Layouts verwenden
 - die Bauteile so groß wie möglich machen (Problem: Kapazitäten, Fläche,..)
 - identische Umgebungen zeichnen ('Dummy Strukturen' = 'etch guards')
 - die Bauteile in der gleichen Orientierung zeichnen (wg. z.B. Piezoresistivität)
 - Bauteile gleichsinnig vom Strom durchlaufen lassen
 - Die Bauteile so nahe wie möglich anordnen, evtl. 'ineinander' legen (s. Beispiel)
 - 'common centroid' Geometrie verwenden, um Gradienten abzufangen (Schwerpunkte zusammengehöriger Bauteile fallen zusammen, s. Beispiel)
 - Temperaturgradienten vermeiden
 - Bauteile nicht an den Rand von Chips legen (Spannungen im Silizium)
 - (Keine Kontakte etc. auf Bauelemente legen, Strukturen in oberen Lagen gleich machen.)
- Sollen Bauteile feste Verhältnisse haben, so sollte man Vielfache von Einheitsbauelementen benutzen
 - Wo das nicht möglich ist, sollte das Verhältnis Rand / Fläche konstant gehalten werden.

Kondensatoren – Matching 1

- Problem: Durch Verschiebung der 2 Plattenlagen kann es zu Toleranzen kommen.



- Lösung: ‚Dummy-Finger‘

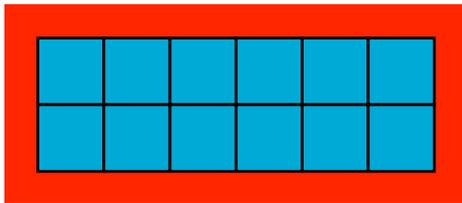


- Zum besseren Matching umgibt man das Layout wieder mit Dummy-Strukturen.

Kondensatoren – Matching 2

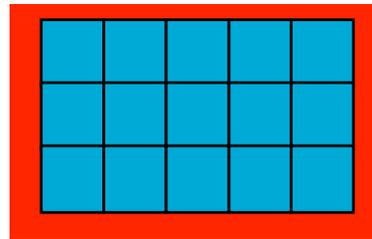
- Verhältnis Fläche/Rand (Area/Periphery) gleich halten:
 - Die Kapazität bei Fläche A und Rand P ist: $C = A \cdot C_{area} + P \cdot C_{fringe} = A \cdot (C_{area} + C_{fringe} \cdot P / A)$
 - Ein **beliebiges Vielfaches** $k \cdot C$ erhält man daher, wenn $A \rightarrow k \cdot A$ unter **Beibehaltung** von P/A gesetzt wird
 - Die einfache lineare Skalierung der Struktur um \sqrt{k} funktioniert nicht, da dann $A \rightarrow k \cdot A$, aber $P \rightarrow 4 \cdot \sqrt{k} \cdot P$ wird.

Gegeben:
,Einheitskapazität'



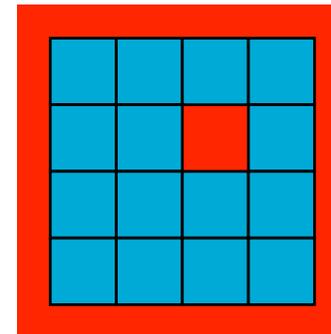
A = 12 Flächeneinheiten (FE)
 $P/A = 16LE / 12FE = 4/3$ LE/FE

Gesucht z.B.:
Layout f. 25% mehr Cap.
i.e. $12 \times 5/4 = 15$ FE
mit GLEICHEM P/A



A = 15 FE
 $P/A = 16/15 = 4/5$ LE/FE ☹️

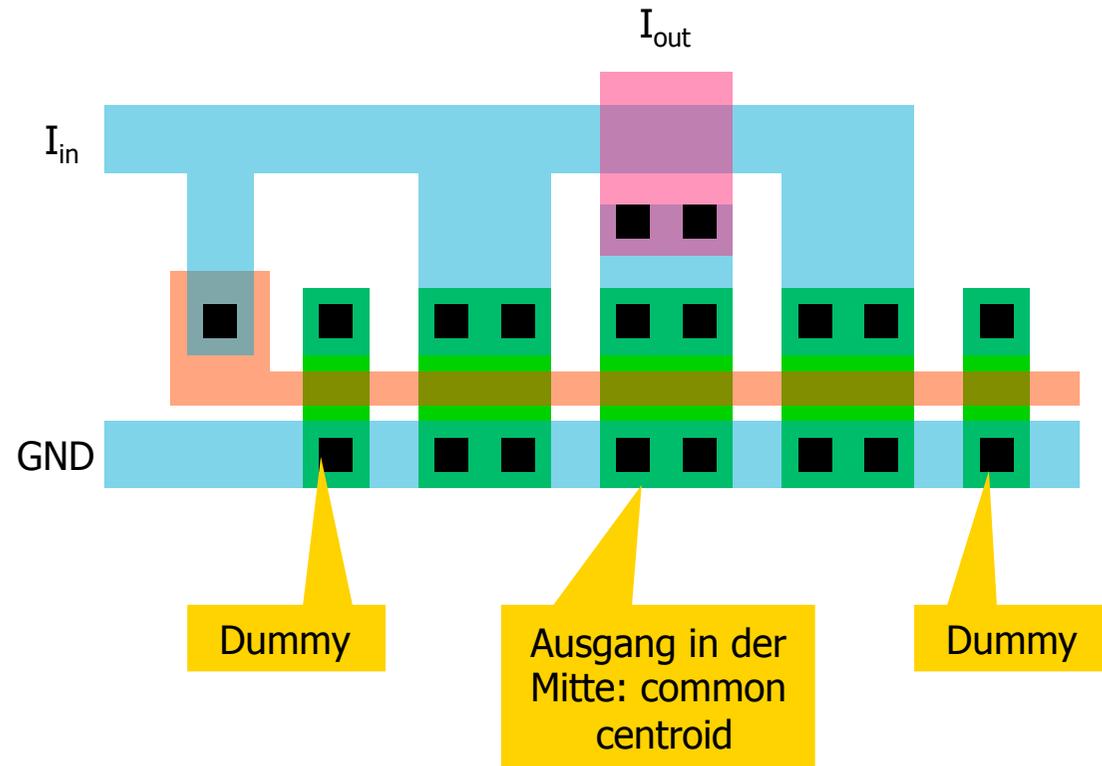
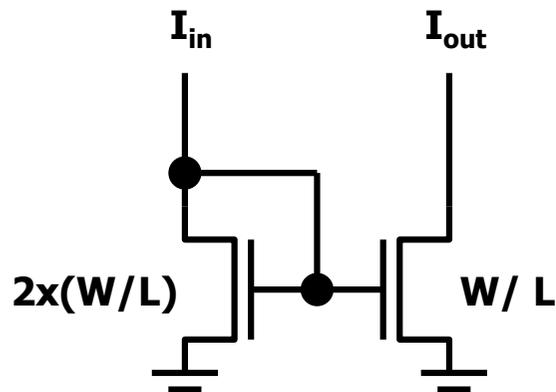
Lösung:
Kapazität mit Loch



A = 15 Flächeneinheiten
 $P/A = 20LE / 15FE = 4/3$ LE/FE

Matching: Transistoren

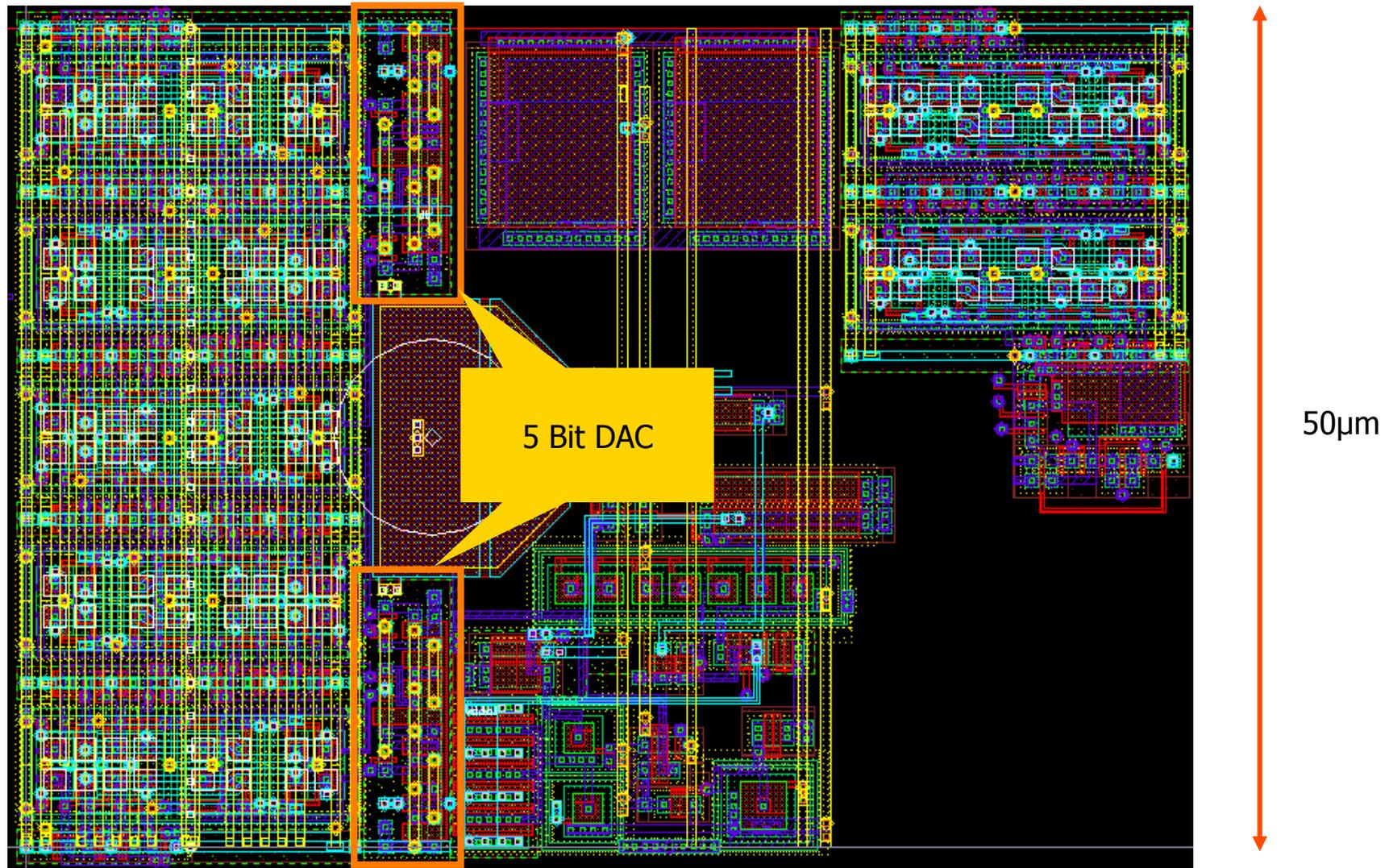
- Wenn exakte Vielfache benötigt werden: Transistoren duplizieren, NICHT Länge oder Breite verändern!
- Beispiel: Präziser Stromspiegel 2:1:
 - keine exotischen Formen
 - gleiche Transistorgeometrie
 - gleiche Stromrichtung
 - gleiche Umgebung (Dummy Strukturen)
 - common centroid zur Elimination von Gradienten



- Präzision ist nicht immer erforderlich !!!

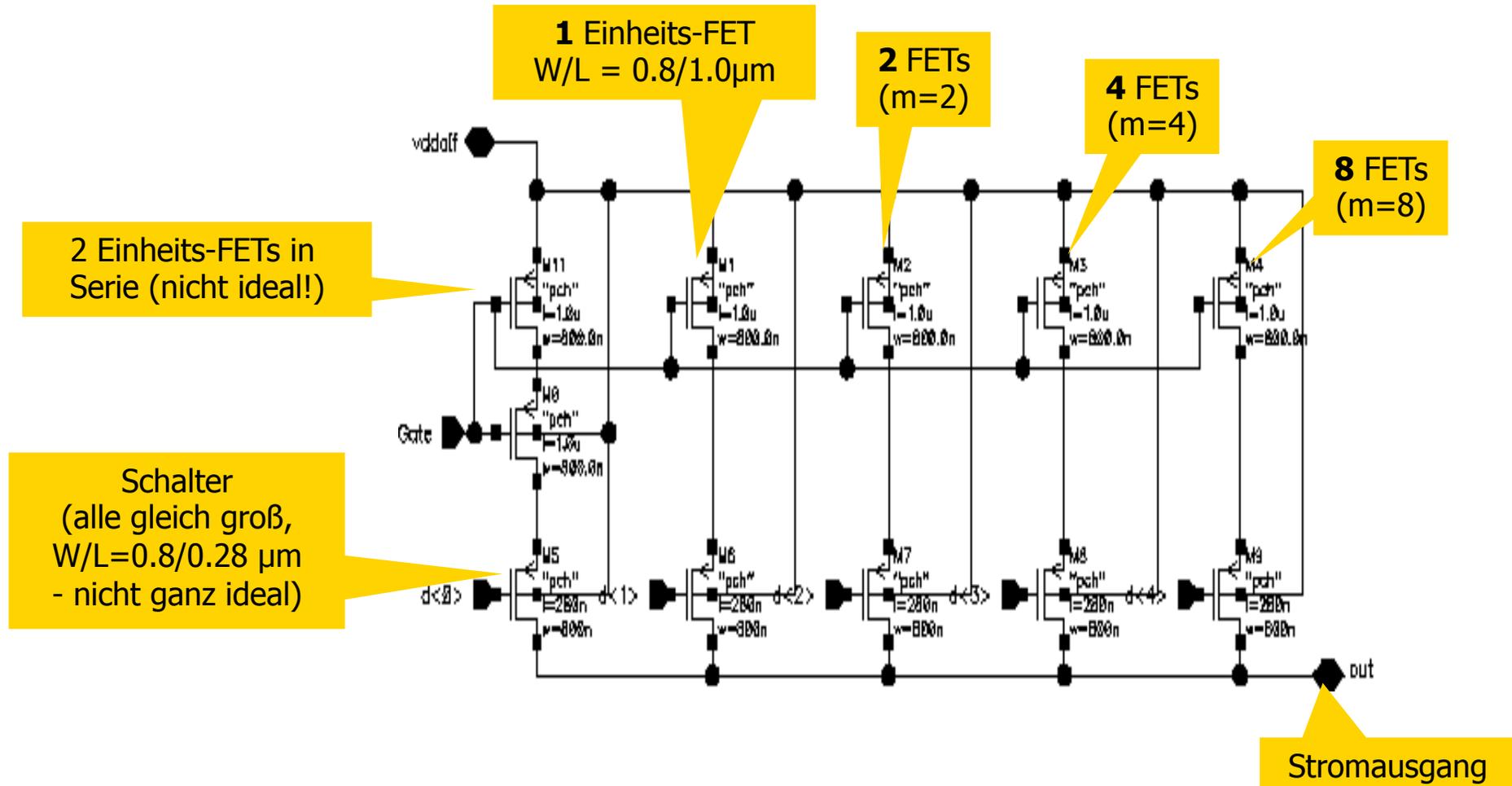
Beispiel aus der Praxis: 5 Bit DAC

- 5 Bit DAC in einem Pixelchip in 0.25 μm Technologie. 2 x 2880 Stück auf einem Chip



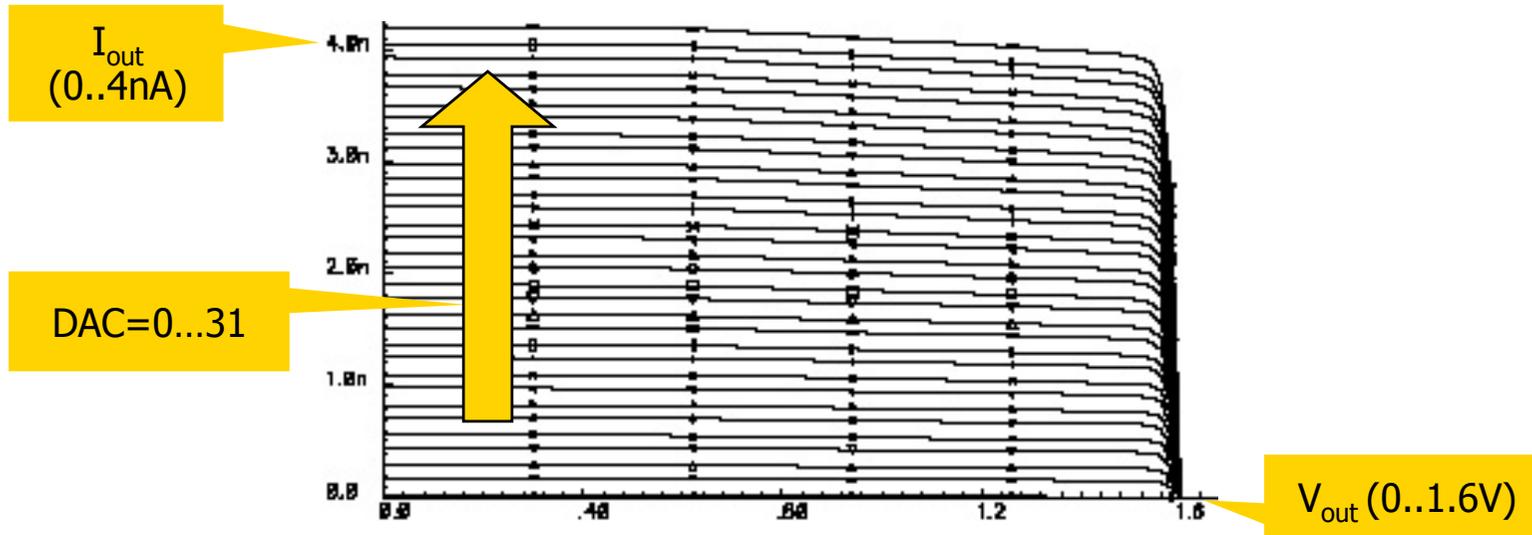
Schaltung

- Sehr einfach: skalierte Stromquellen werden zu- und abgeschaltet

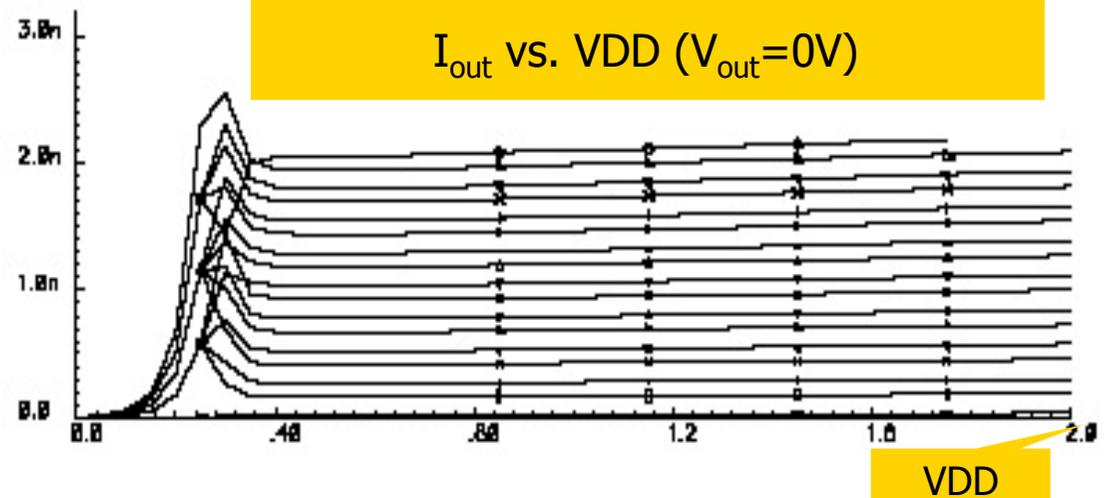


Simulation ist sehr linear

Ausgangskennlinie (VDD=1.6V)

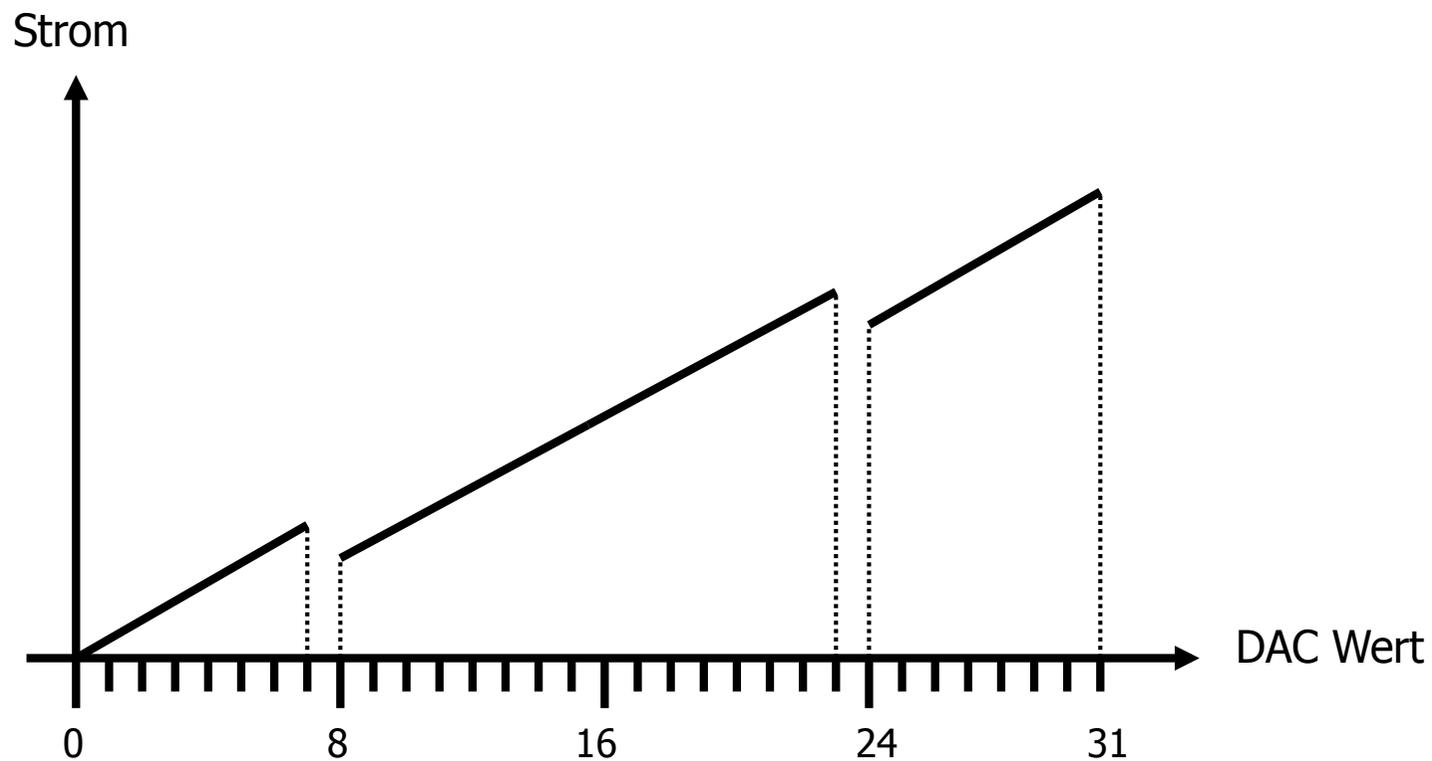


I_{out} vs. VDD ($V_{out}=0V$)



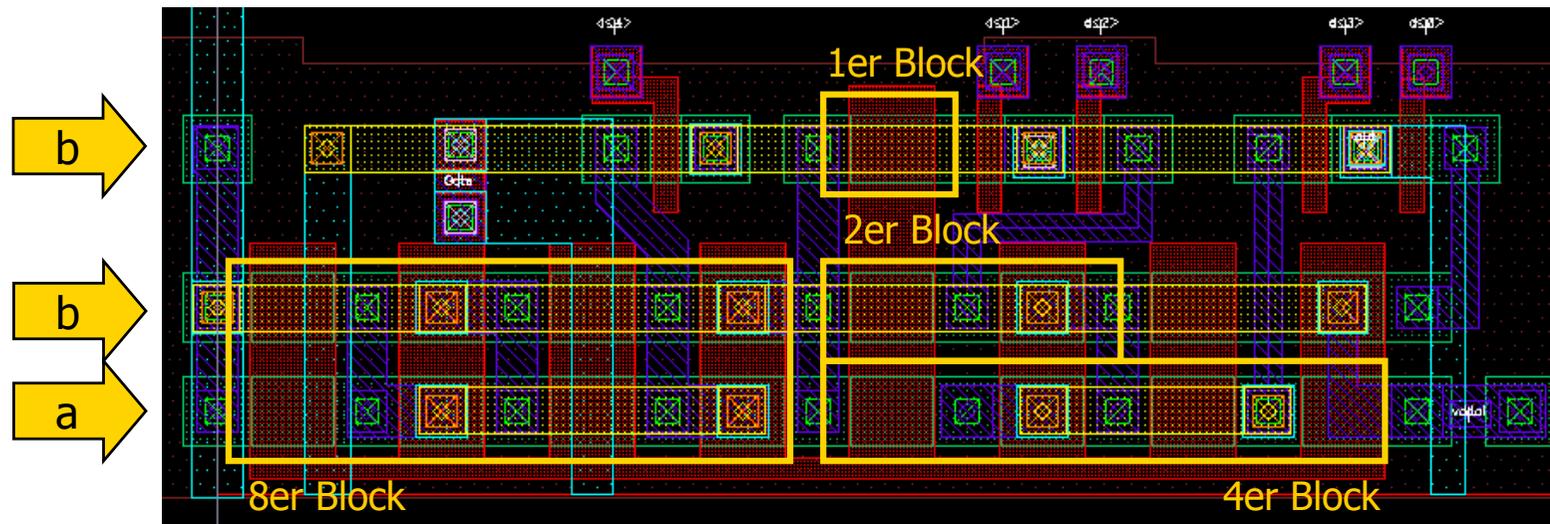
Messung, schematisch

- Große Stufen bei $7 \Rightarrow 8$ and $23 \Rightarrow 24$.
- Kein Problem bei $15 \Rightarrow 16$ (das ist normalerweise der kritische Punkt!)

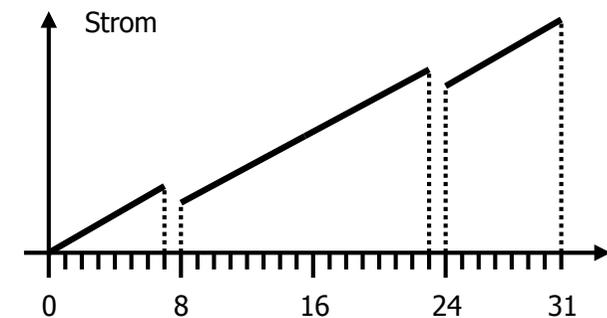


Mögliche Erklärung

- Erklärungsversuch: Obere und untere Transistorreihe haben unterschiedlichen Strom (Evtl. wegen der Poly1 Leitung unterhalb der unteren Transistorreihe)



- Test: Annahme: FETs in unterer Reihe erzeugen den Strom **a**, FETs in oberer Reihe den Strom **b**. ($a > b$)
 - Schritte < 7 : je $0.5b$
 - Schritt bei $7 \Rightarrow 8$: $3.5b \Rightarrow 4a$ Fehler: $4 \cdot (a - b)$
 - Schritt bei $15 \Rightarrow 16$: $3.5b + 4a \Rightarrow 4b + 4a$ kein Fehler
 - Schritt bei $23 \Rightarrow 24$: $7.5b + 4a \Rightarrow 8b + 4a$ Fehler: $4 \cdot (a - b)$



Verschiedenes

Spannungsabfälle, Elektromigration

Spannungsabfälle

- Hohe Ströme führen durch den Leitungswiderstand zu Spannungsabfällen ('IR-Drops').
Hiervon sind insbesondere Versorgungsspannungen (statisch) und Clock-Treiber betroffen (dynamisch)
 - Breite Leitungen benutzen
 - Metallagen mit niedrigem Widerstand benutzen
 - Mehrere Lagen benutzen
 - Kontakte vervielfältigen

Elektromigration

- bei hohen DC Strömen wandern die Metallatome und erodieren mit der Zeit das Metall
- Durch diese Elektromigration können Leitungen unterbrochen werden oder (durch laterale Ablagerung und die Bildung von nadelförmigen 'Whiskers') Kurzschlüsse entstehen
- Dies kann langfristig zum Ausfall des Chips führen.
- Die MTF (Mean Time to Failure) ist proportional zu $1/I^2$ und hängt exponentiell von der Temperatur ab.
Durch Test bei hoher Temperatur kann man daher die MTF ermitteln ('accelerated aging').
- Die Stromdichte muß überall unter einem (hoffentlich vom Hersteller vorgegebenen) Limit bleiben
Literaturwert: max. 5×10^5 A/cm² für Cu-dotiertes Aluminium bei 85°C (0.5-4% Cu erhöht Al-Haltbarkeit)
Faustregel: max. 1.5mA pro μm Bahnbreite
- Leiterbahnen, die über Oxid-Stufen laufen, sind anfälliger (da dort meist etwas dünner)

Skineffekt

- Bei niedrigen Signalfrequenzen fließt der Strom im gesamten Volumen eines Leiters
- Bei **sehr hohen Frequenzen** (GHz) werden die Ladungsträger durch das von ihnen selbst erzeugte Magnetfeld aus dem Inneren des Leiters verdrängt.
- Stromfluss findet nur noch in einer dünnen ‚Haut‘ - Schicht (skin) an der Oberfläche statt.
- Die Anzahl Ladungsträger nimmt exponentiell mit der Eindringtiefe ab.
- In der **Skin Tiefe** δ („skin depth“) ist die Stromdichte auf $1/e$ abgefallen.



niedrige Frequenz



hohe Frequenz

$$\delta = \text{sqrt} (1 / \pi f \mu \sigma) = \text{sqrt} (\rho / \pi f \mu)$$

mit

f = Signalfrequenz,

μ = (magnetische) Permeabilität = $4\pi \cdot 10^{-7}$ H/m,

σ = Leitfähigkeit,

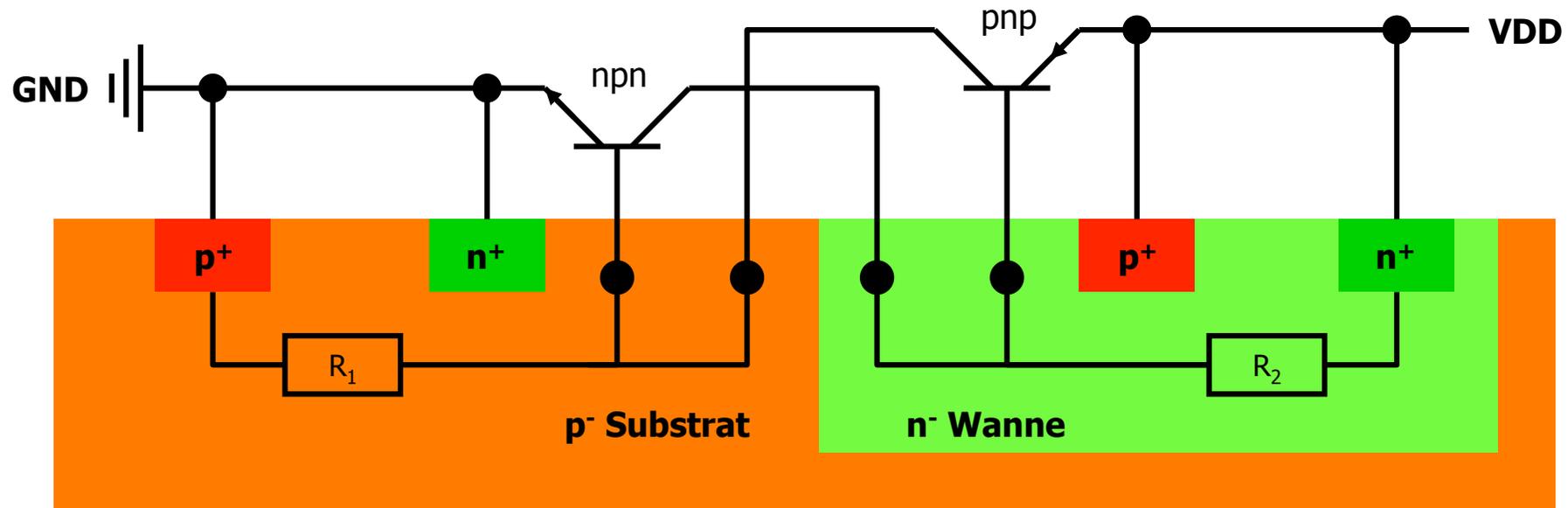
ρ = $1/\sigma$ = spez. Widerstand

- Beispiele:
 - Für Aluminium ($\rho=2.65\mu\Omega\text{cm}$) bei 1 GHz ist $\delta=2.6 \mu\text{m}$
 - Für Kupfer ($\rho=1.67\mu\Omega\text{cm}$) bei 10 GHz ist $\delta=0.7 \mu\text{m}$

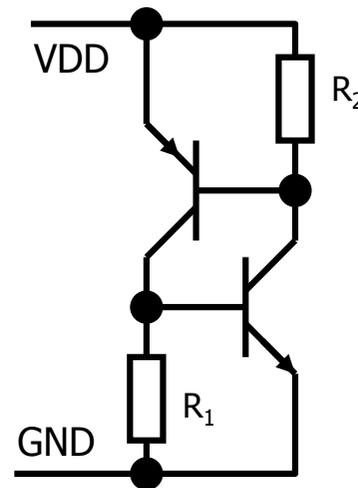
- Ergebnis: Bei Leiterbahndicken von μm spielt der Skineffekt in Chips (noch) **keine große Rolle**
- Herleitung z.B.: <http://scienceworld.wolfram.com/physics/SkinDepth.html>

Latchup

- Durch die verschieden implantierten Bereiche gibt es parasitäre Strukturen aus npn und pnp Transistoren

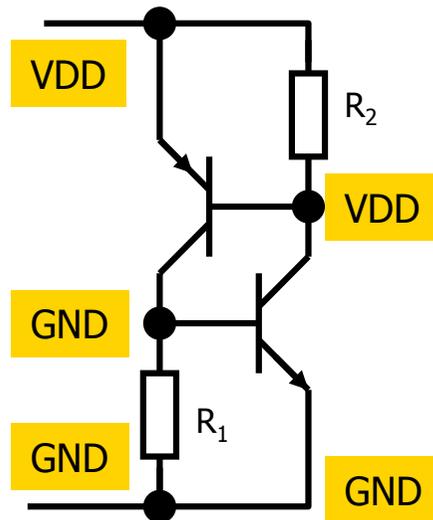


Äquivalentes Schaltbild:

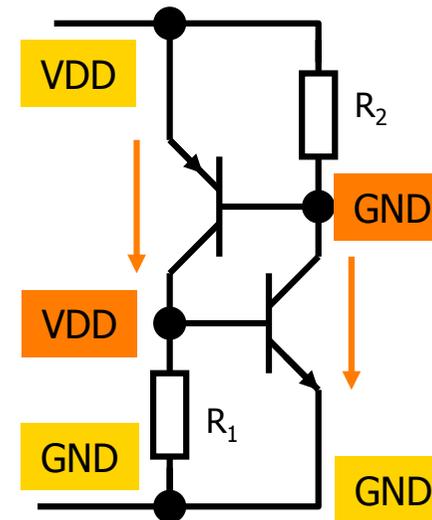


Latchup

- Die äquivalente Schaltung kann 'gezündet' werden, wenn eine BE-Diode in Vorwärtsrichtung leitet
- Dies kann durch kurze Spannungsspitzen passieren
- Die npnp-Struktur schaltet dann ein (Thyristor, SCR=Silicon Controlled Rectifier) und BLEIBT angeschaltet
- Der hohe Querstrom zwischen VDD und GND kann die Schaltung zerstören



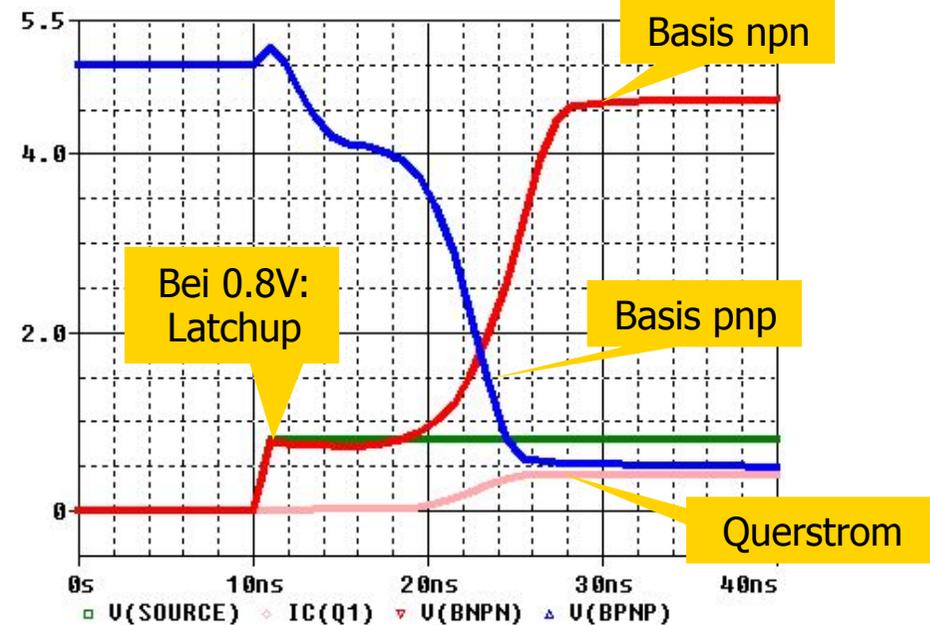
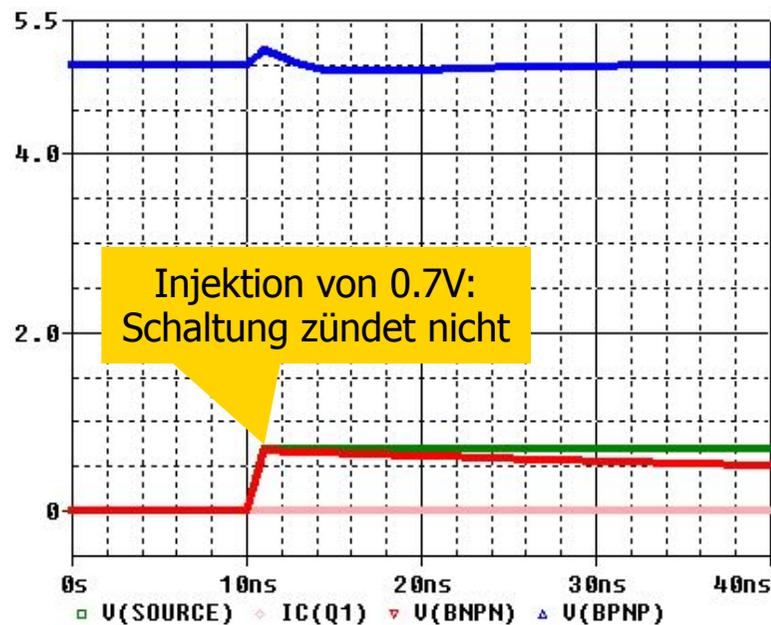
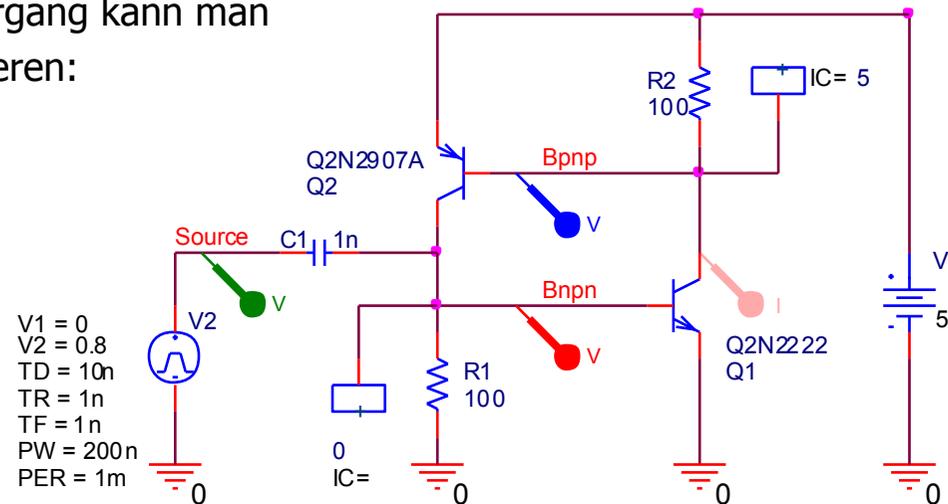
Normalzustand:
Beide Transistoren sind OFF



Gezündeter Zustand (vereinfacht):
Beide Transistoren sind ON

Latchup Simulation mit SPICE

- Den beschriebenen Vorgang kann man leicht mit SPICE illustrieren:



Latchup Vermeidung

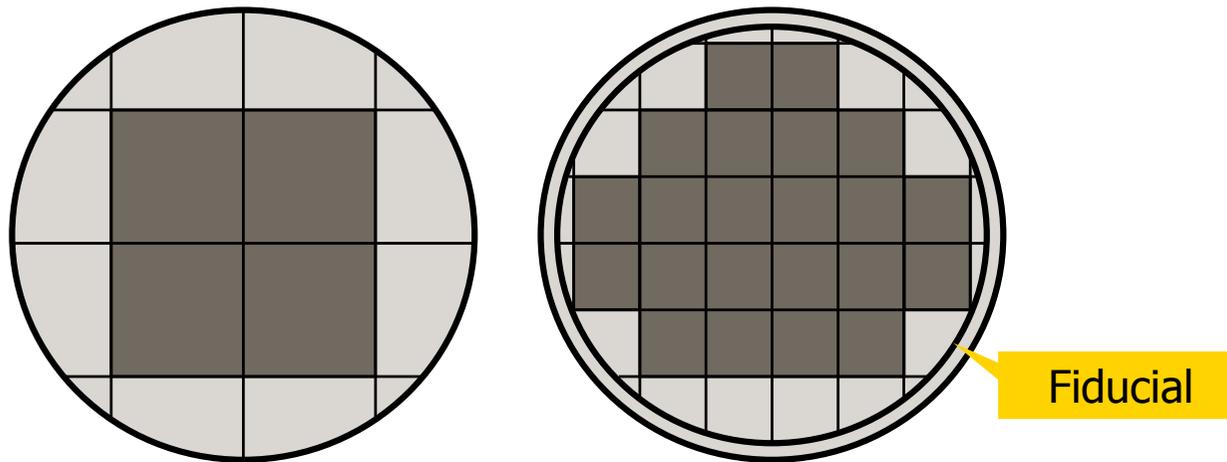
- Die Schleifenverstärkung muß klein sein. Dies kann durch geeignete Technologieschritte erreicht werden
- Die Dioden dürfen nicht in Leitung kommen:
 - Widerstände müssen klein sein
 - **Substrat/Wannenkontakte nahe bei den Transistoren**
 - Niederohmige Anbindung der Kontakte an Versorgungen und Transistoren (Metall!)
 - **Jede Wanne muß Kontakte bekommen!**
 - Abstand der Transistoren vergrößern
 - Guard-Ringe benutzen
- Besondere Vorsicht ist bei Schaltung mit hohen Transientenströmen (Buffer, IO Pads) geboten

Ausbeute (Yield)

Ausbeute (Yield)

$$\text{Yield} = \frac{\text{Anzahl } \mathbf{gute} \text{ Chips ('die')} \text{ auf dem Wafer}}{\text{Anzahl alle Chips auf dem Wafer}}$$

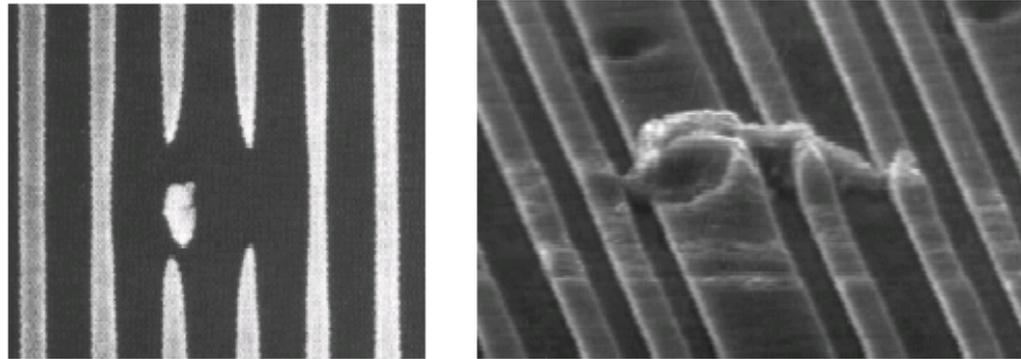
- Ein gewisser am Rand ist unbrauchbar (Prozessierung dort schlecht, Verletzungen durch Anfassen des Wafers,...). Es werden daher nur die Chips innerhalb des brauchbaren Bereichs ('fiducial area,') gezählt.



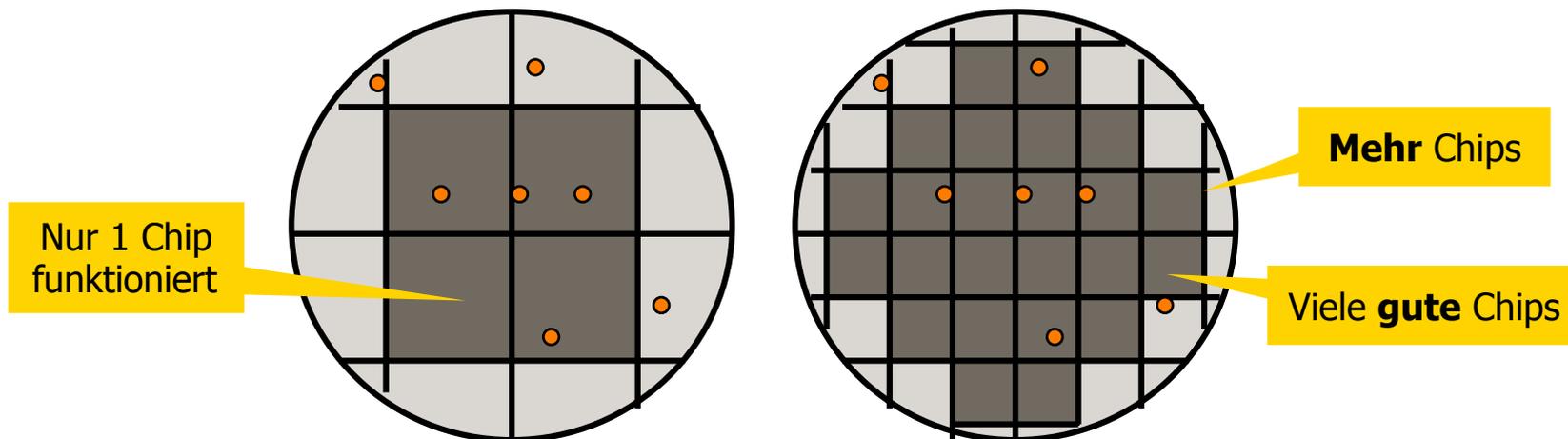
- Partikel größer als \sim min. Strukturbreite/3 können (müssen aber nicht, $p=0.2-0.4$) zu Totalausfall führen
- Beispiel $0.5\mu\text{m}$ Prozeß, 1cm^2 Chip: Fußballfeld – Stecknadelkopf
- Nur 4-5 Ebenen sind kritisch (Poly, Implantationen, unteres Metall), dann werden die Strukturen größer.
- Eine Gesamtausbeute (Wafer \Rightarrow fertig verpackte Bauteile) von 70% ist gut.

Defekte

- Einfachste Defekte: open / short



- Die Wahrscheinlichkeit für den Ausfall eines Chips ist um so höher, je größer der Chip ist.
- Gleichzeitig gibt es bei großen Chips sehr viel weniger Chips auf dem Wafer



⇒ **Die Ausbeute nimmt mit der Größe der Chips sehr stark ab!**