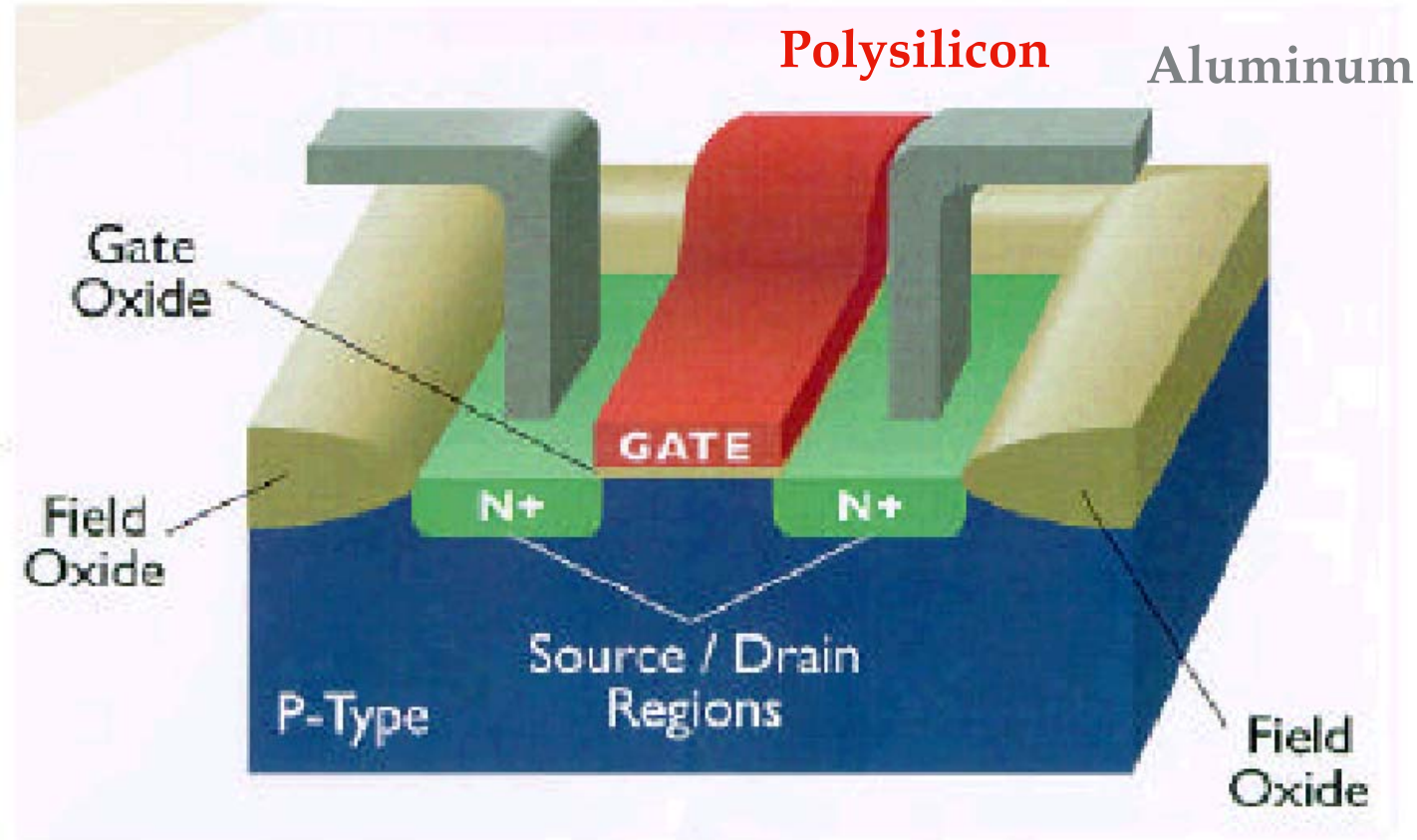
EEC 116 Lecture #2: MOSFET Basics Layout Introduction CMOS Inverters

Rajeevan Amirtharajah Bevan Baas
University of California, Davis
Jeff Parkhurst
Intel Corporation

Outline

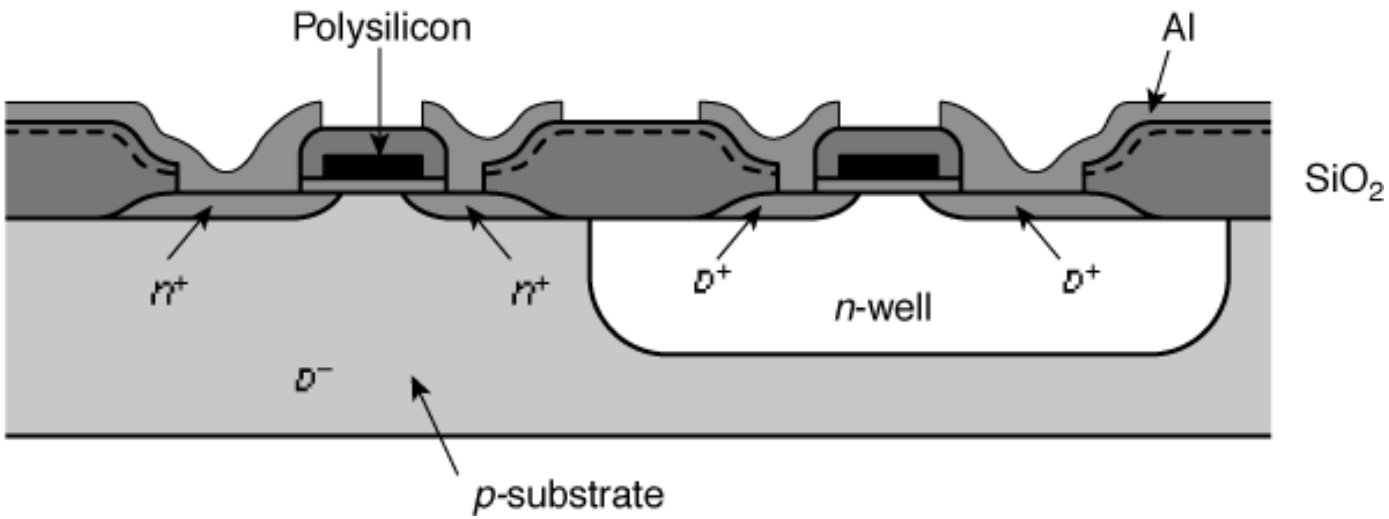
- **Finish Lecture 1 Slides**
- **Switch Example**
- **MOSFET Structure**
- **MOSFET Regimes of Operation**
- **Scaling**
- **Parasitic Capacitances**

3D Perspective



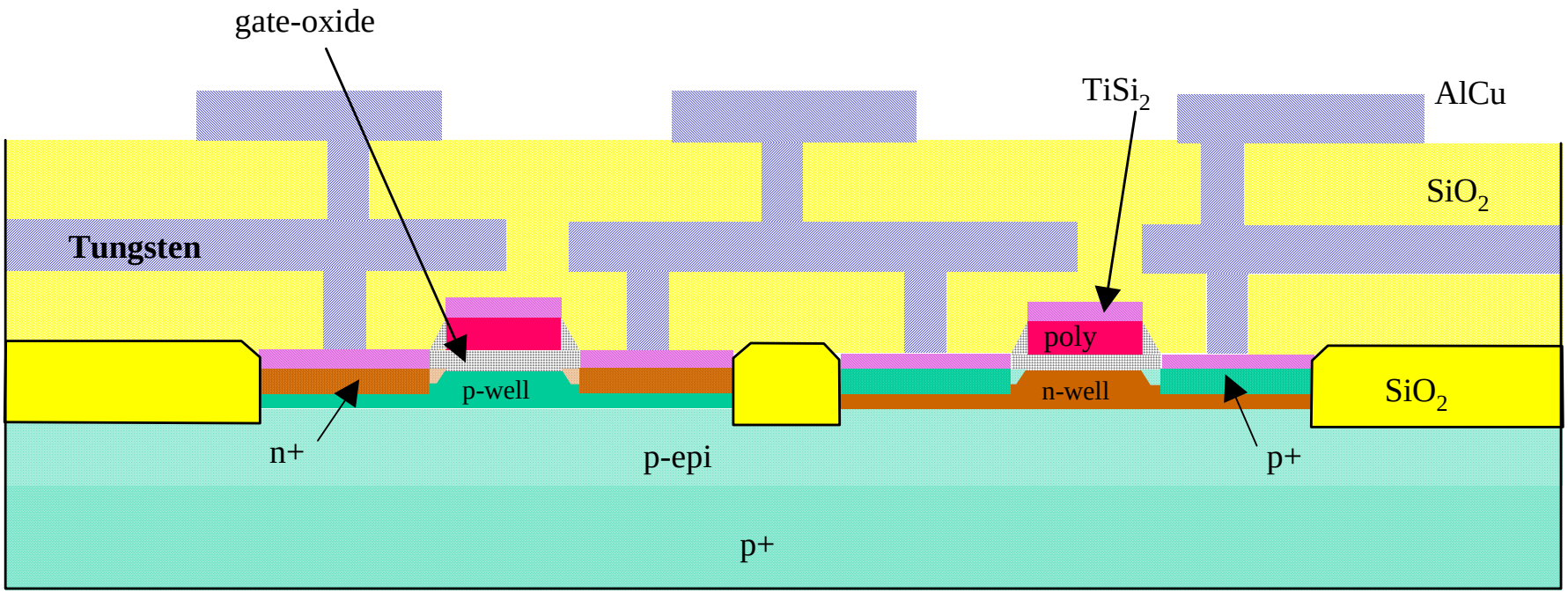
Source: Digital Integrated Circuits, 2nd ©

CMOS Process



Source: Digital Integrated Circuits, 2nd ©

A Modern CMOS Process

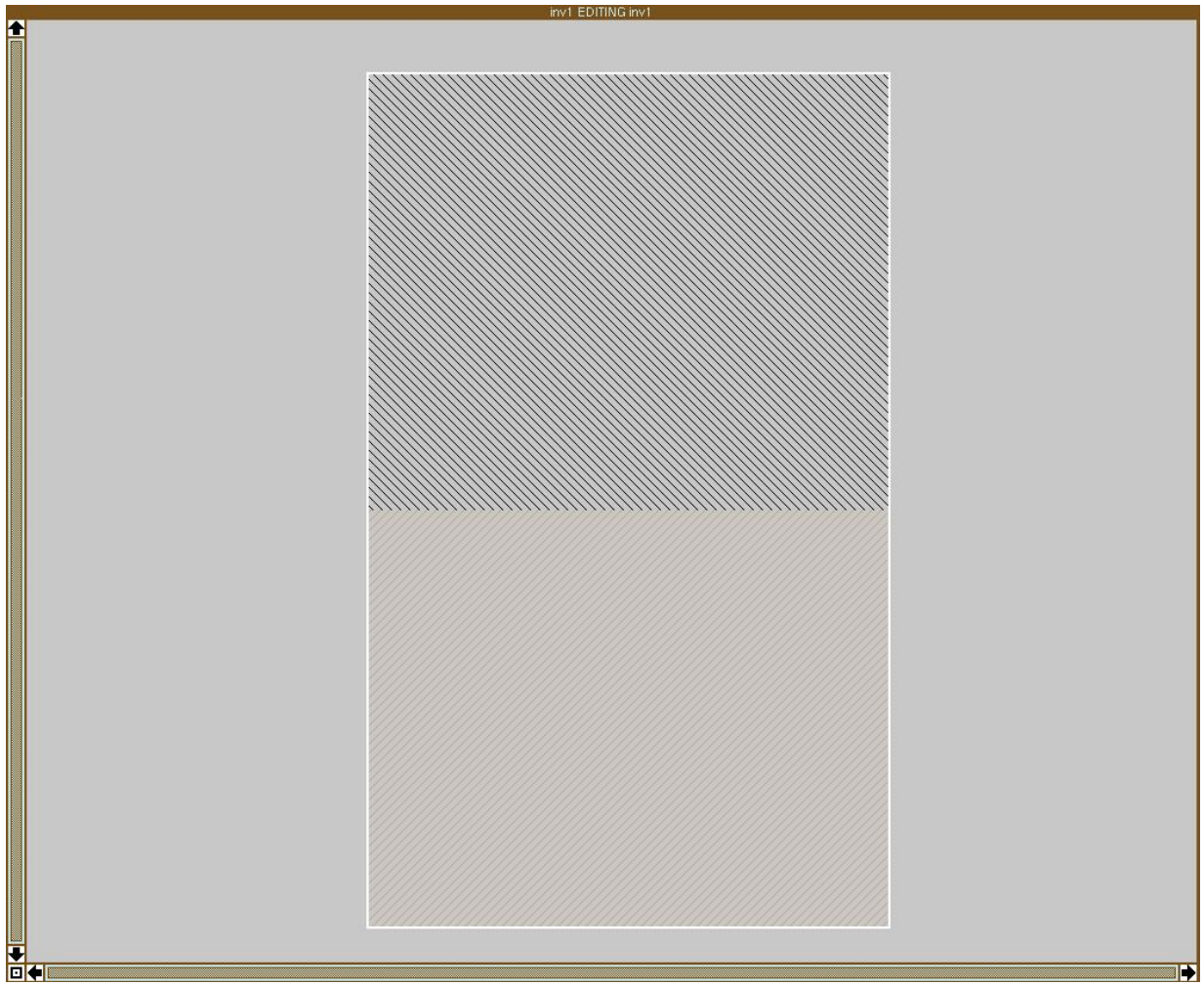


Dual-Well Trench-Isolated CMOS Process

Source: Digital Integrated Circuits, 2nd ©

nwell and pwell

- **The "bodies" of the transistors**



Source: Omar Sattari

ndiffusion and pdiffusion

- **Source and Drain for each transistor**

Source: Omar Sattari

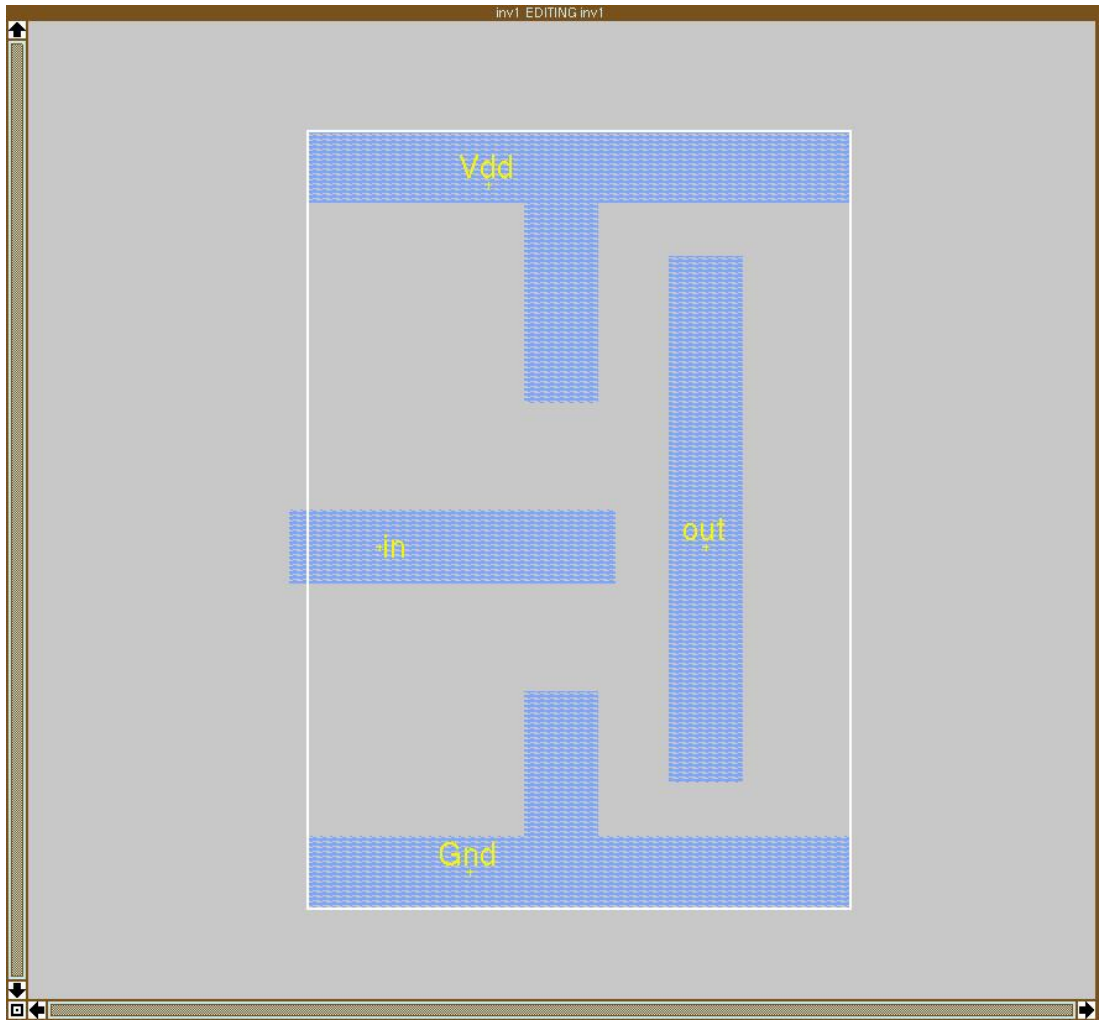
Polysilicon

- **Gate of transistors and for short-distance wiring**

Source: Omar Sattari

Metal1

- **First level of interconnect**



Source: Omar Sattari

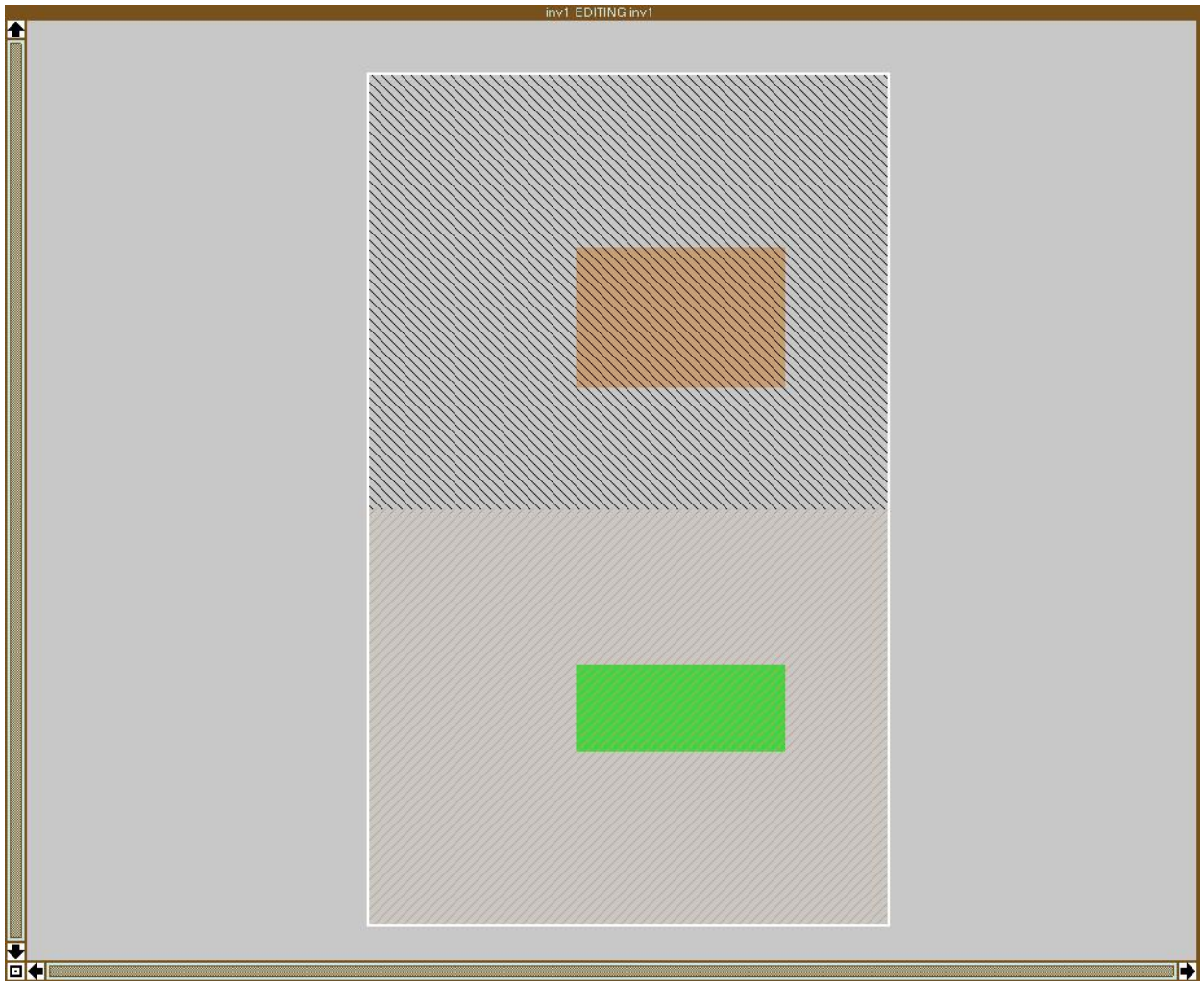
Metal2

- **Second layer of interconnect**

Source: Omar Sattari

Building an Inverter: Starting with Well and Diffusion

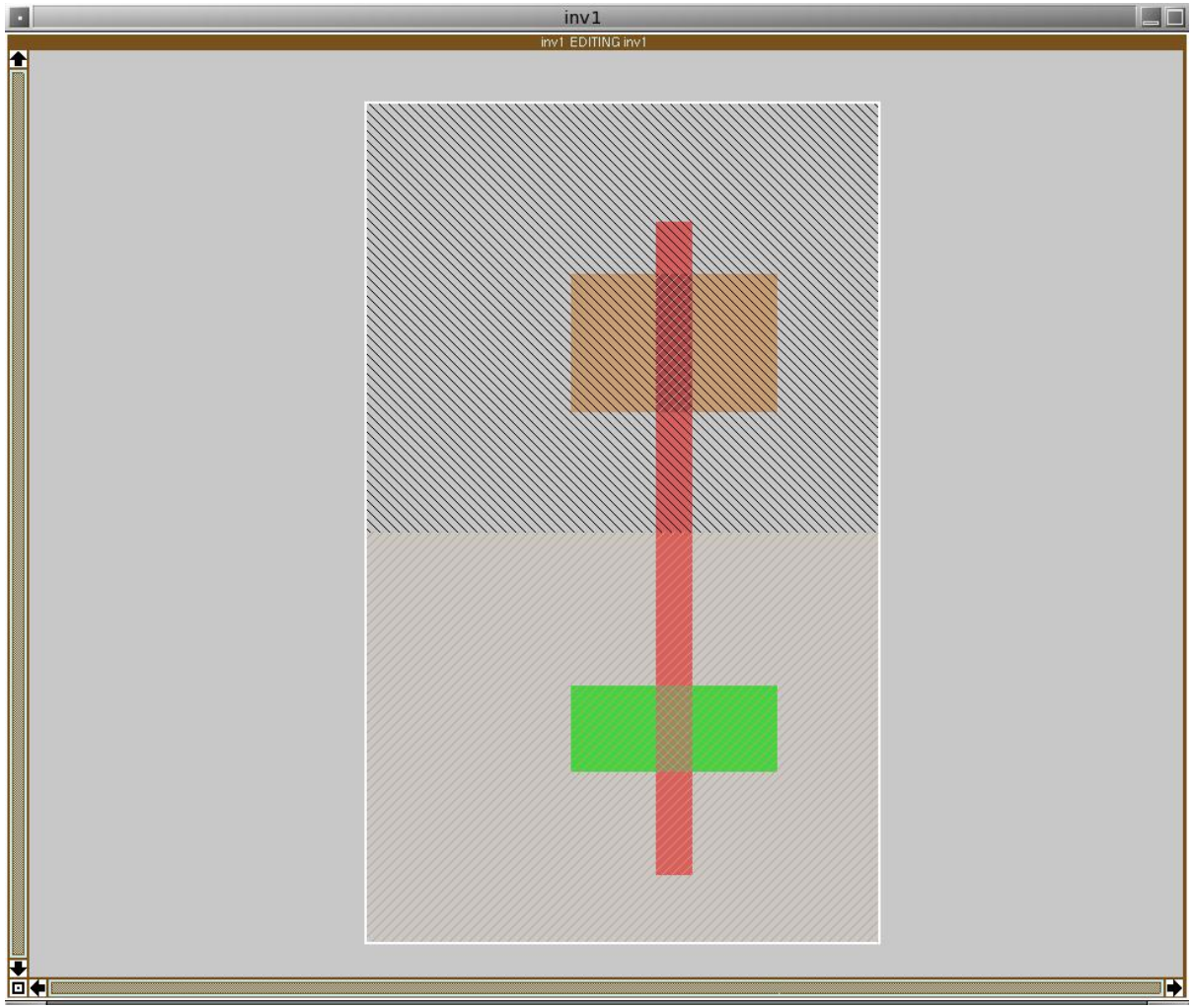
- **Place N-type and P-type diffusions**
  - Convention is to place PMOS on top and NMOS on bottom



Source: Omar Sattari

Transistors

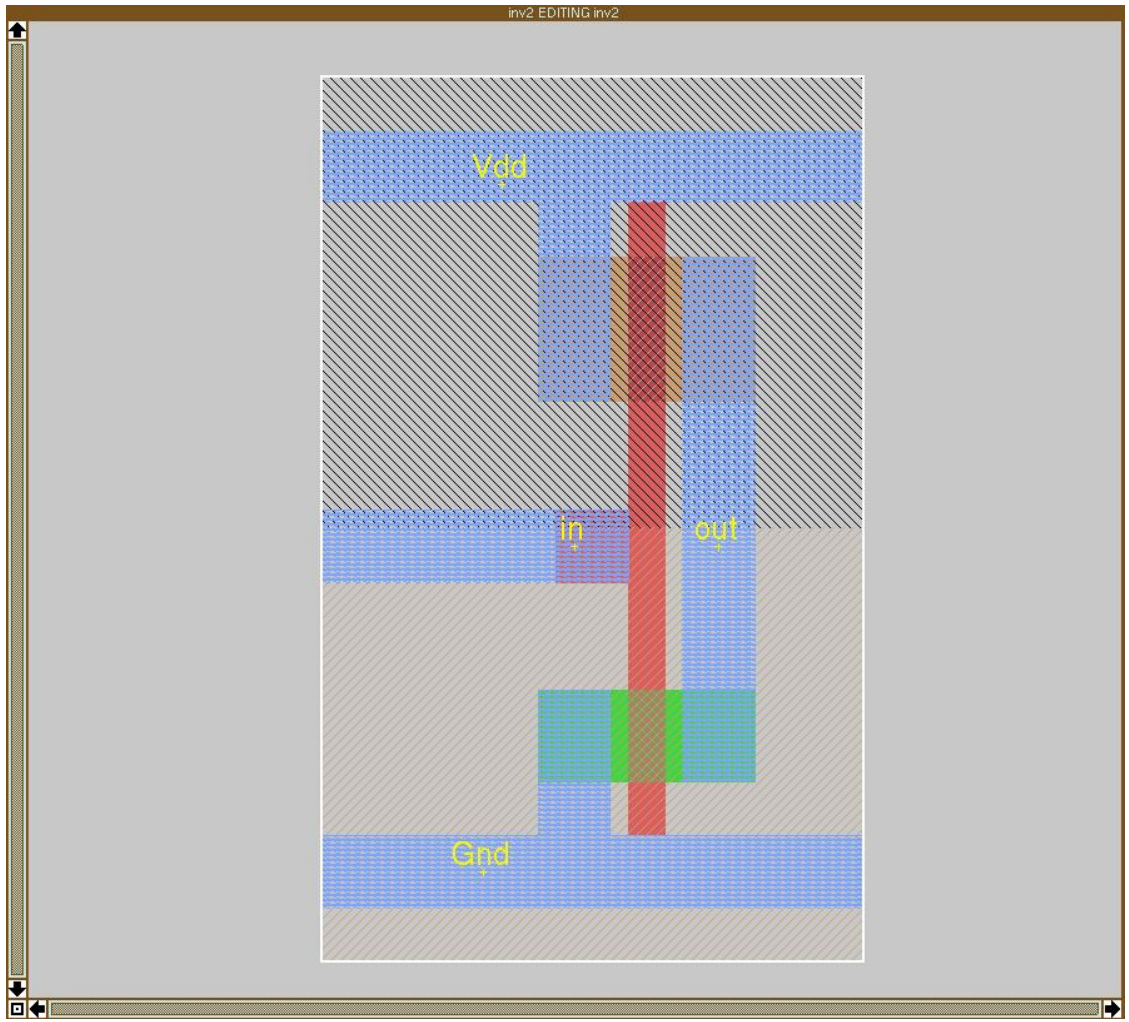
- poly crossing diffusion produces a transistor!
- Common gate here
- PMOS shown on top
- NMOS shown on bottom



Source: Omar Sattari

Metal1

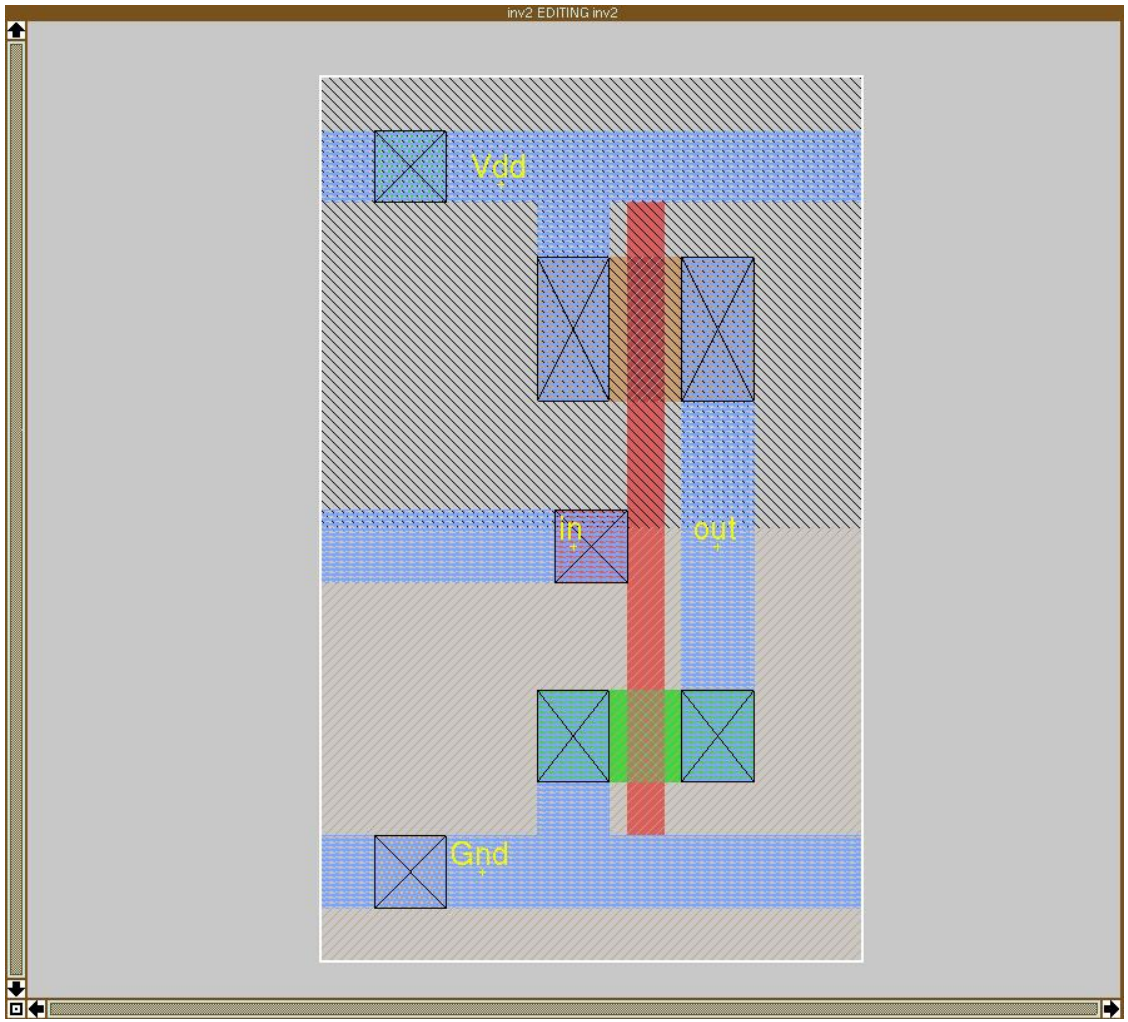
- **metal1 laid down but not yet connected**
- **Use metal for *Vdd* and *Gnd***
- **Labels added**
  - Extremely useful for testing
  - Documents design
  - Use “point” labels, not large area ones
  - Never use *global* labels that end in an “!”



Source: Omar Sattari

Metal1 Contacts

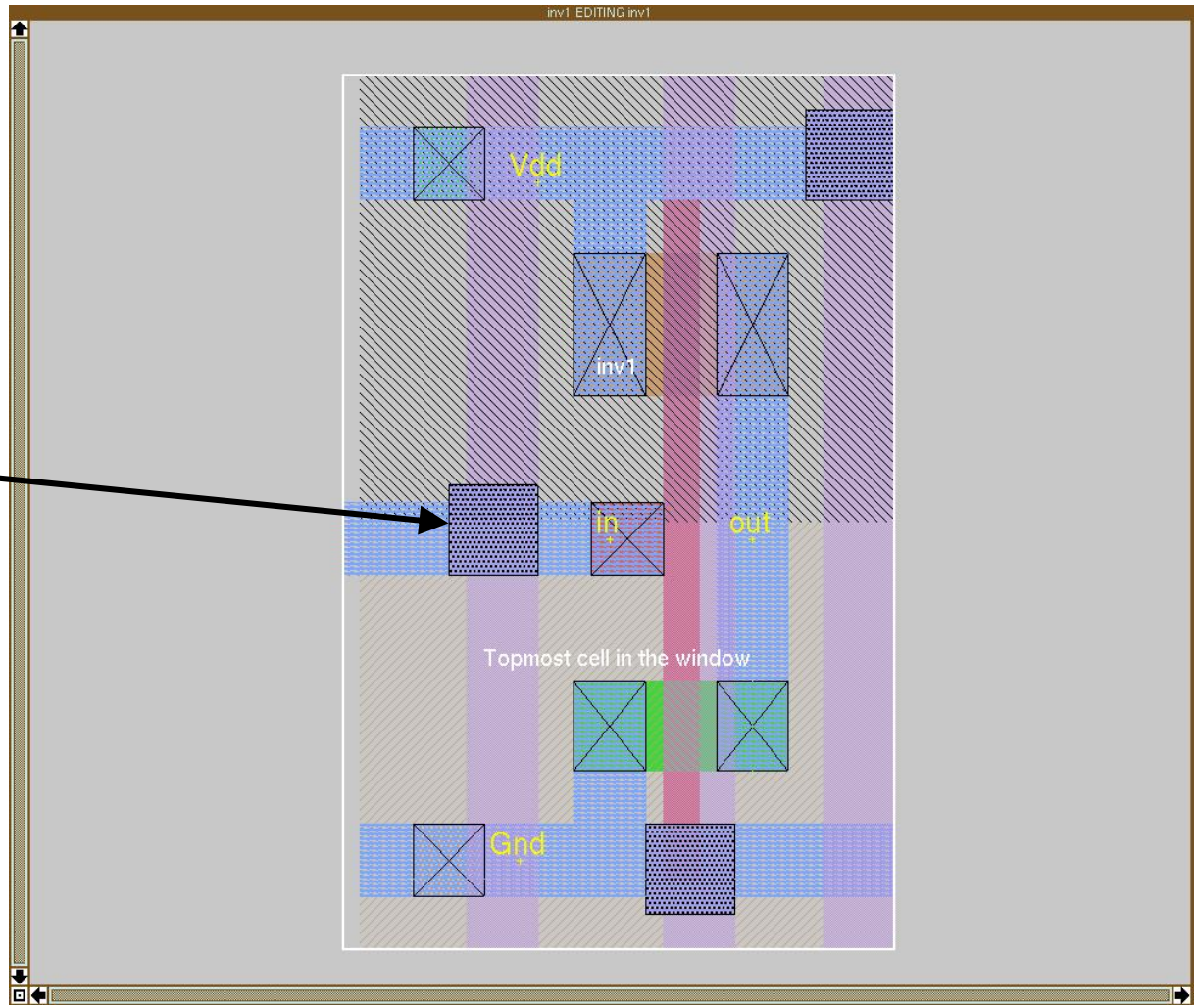
- **Connections now made between metal1 and:**
  - pdiff
  - ndiff
  - poly
  - nwell
  - pwell
- **Each via/contact is a different layer**



Source: Omar Sattari

Metal2

- Use metal2 for longer distance routing
- Routes over the “top” of other circuits shown
- metal2 contacts connect metal1 and metal2 only



Source: Omar Sattari

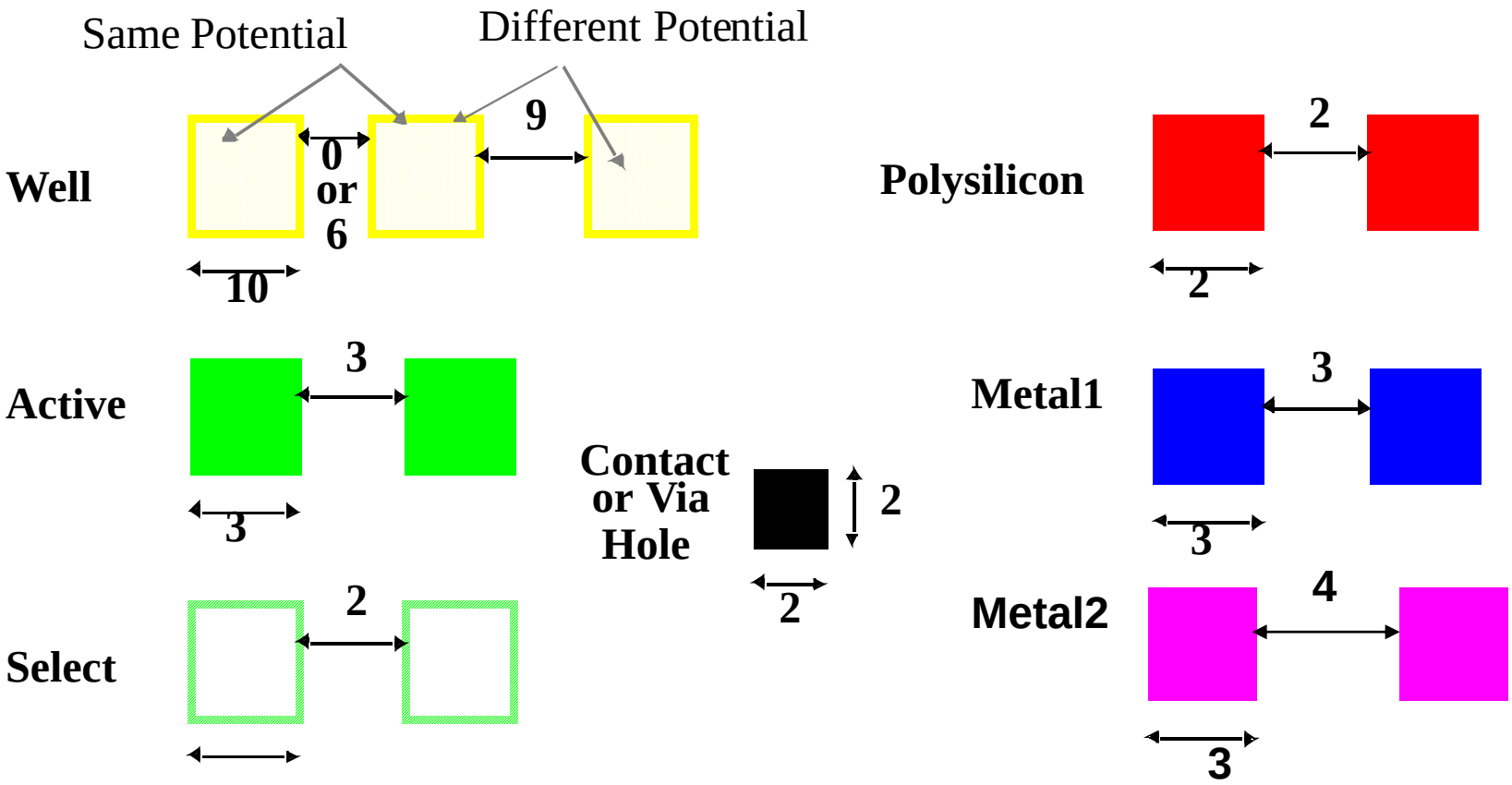
Design Rules

- **Interface between designer and process (CMOS fabrication) engineer**
- **Guidelines for constructing fabrication masks**
- **Units commonly used**
  - Scalable design rules: lambda (λ) parameter (used in magic), or
  - Absolute dimensions (micron rules)
- **Common rule examples:**
  - Minimum width
  - Minimum separation same material
  - Minimum separation different material
- **Look for flashing markers in Cadence that show errors**

Design Rules

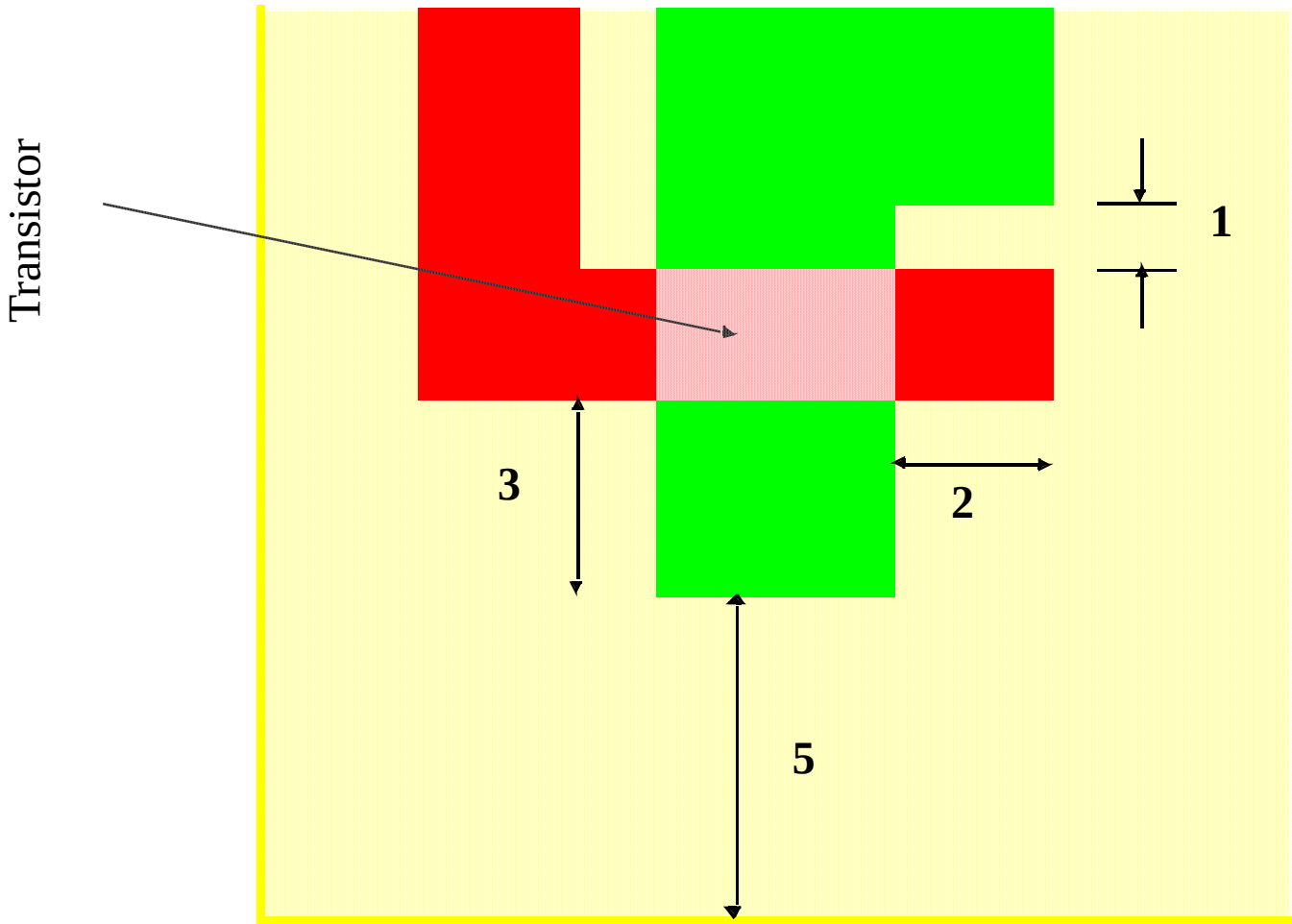
- **Mead and Conway, 1980**
  - "Lambda-based" scalable design rules
  - Allows full-custom designs to be easily reused from technology generation to technology generation
  - Lambda is roughly one half the minimum feature size
    - "1.0 μm technology" -> 1.0 μm min. length, lambda = 0.5 μm
    - "0.5 μm technology" -> 0.5 μm min. length, lambda = 0.25 μm
  - For our class, we are using a 0.18 μm technology so lambda is 0.09 μm
- **See lab handouts for URL to our scalable design rules on the MOSIS website**
- **We are using "SCMOS_DEEP" rules**

Example Intra-Layer Design Rules



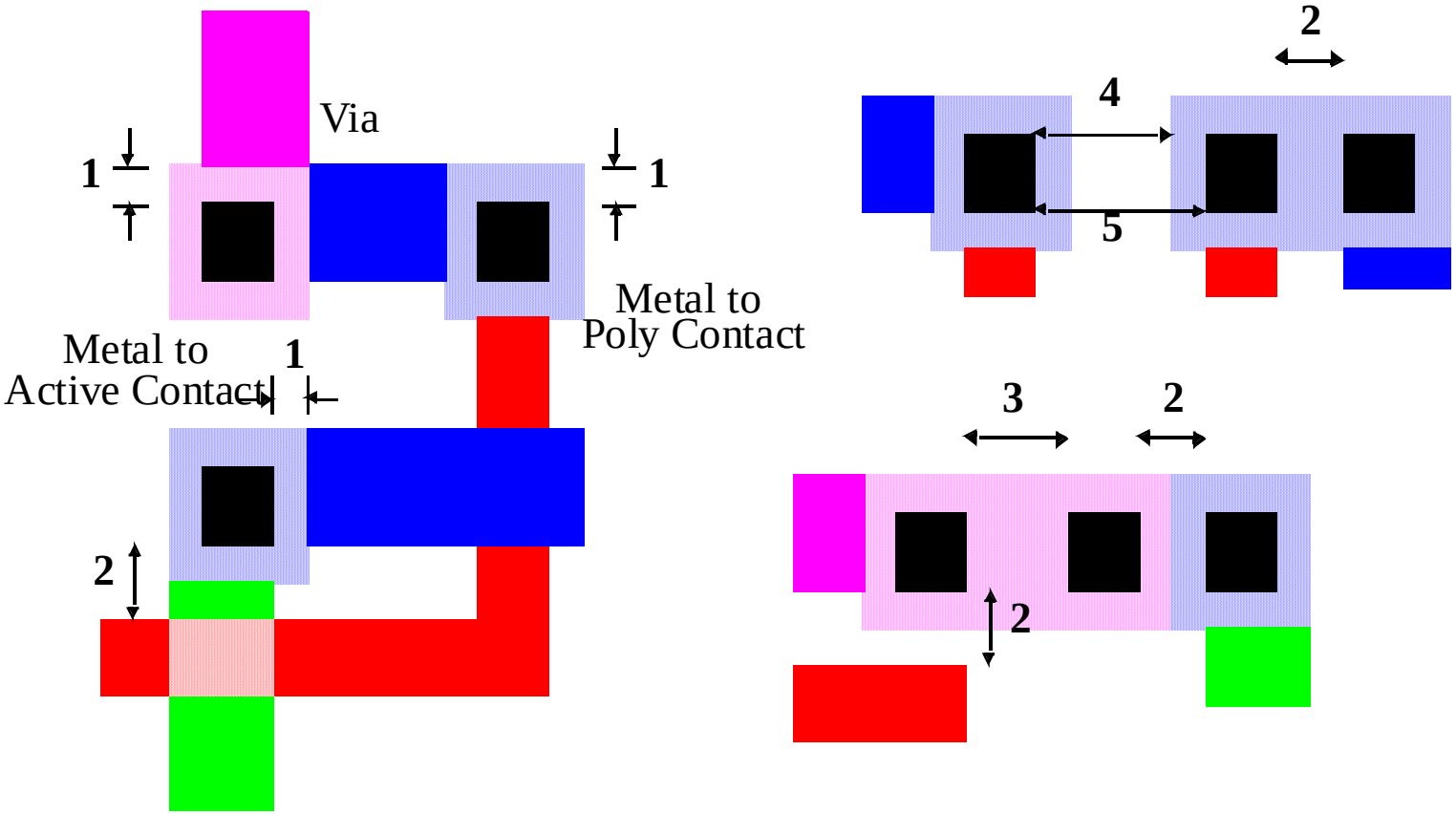
Source: Digital Integrated Circuits, 2nd ©

Example Design Rules: Transistor Layout



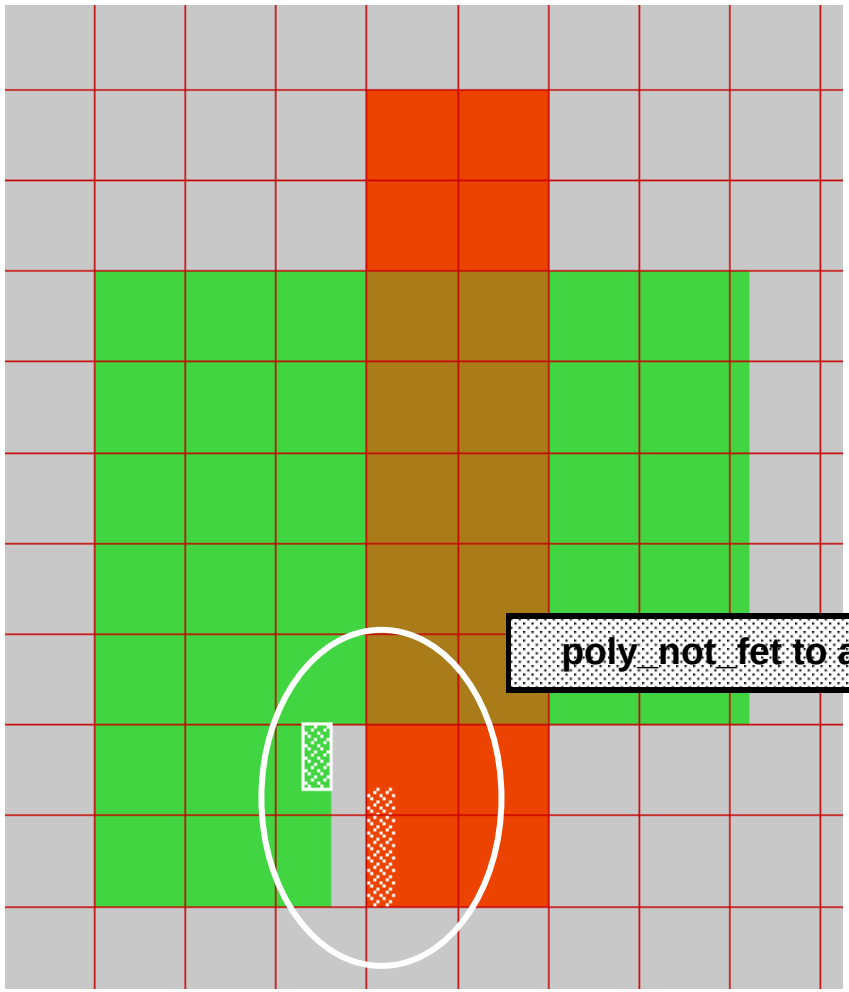
Source: Digital Integrated Circuits, 2nd ©

Example Design Rules: Vias and Contacts



Source: Digital Integrated Circuits, 2nd ©

Design Rule Checker



In magic, white dots appear at the point of a DRC rule violation. Something similar occurs in Cadence.

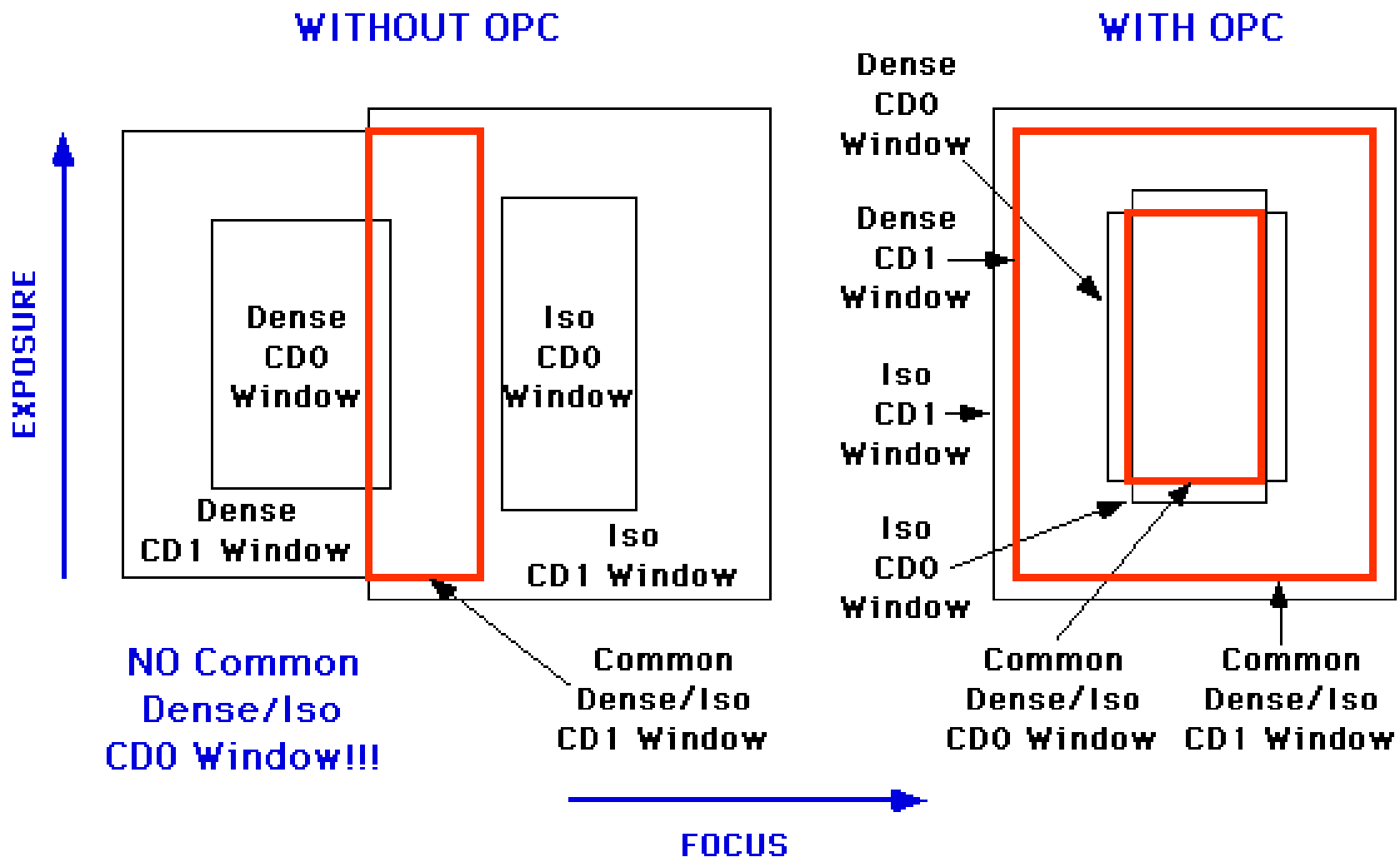
Place a box around white dots and press “y” to see what is causing an error

Source: Digital Integrated Circuits, 2nd ©

Accounting for VDSM Effects

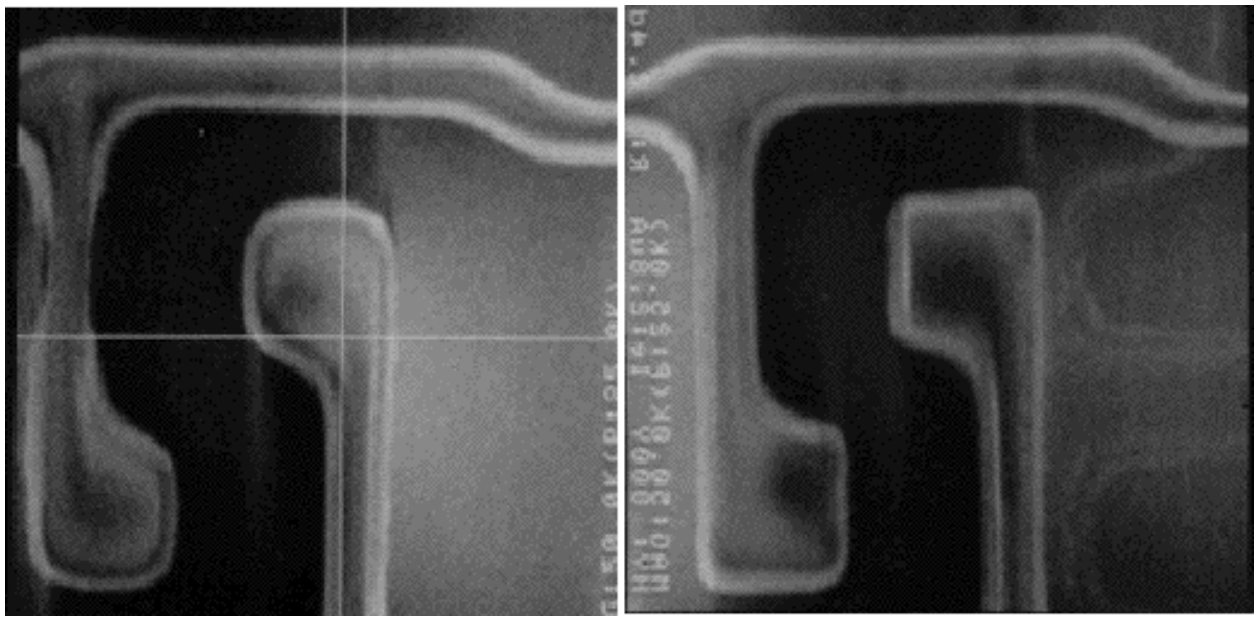
- **VDSM = Very Deep Sub Micron**
  - Effects significant below 0.25 μm (0.18 μm, 130 nm, 90 nm, 65 nm, 45 nm)
- **Compensation made at the mask level**
  - OPC – Optical Proximity Correction
  - Occurs when different mask layers don't align properly
  - Test structures are used to characterize the process
  - Ability to adapt depends on the consistency of the error from process run to process run

Accounting for VDSM Effects: OPC



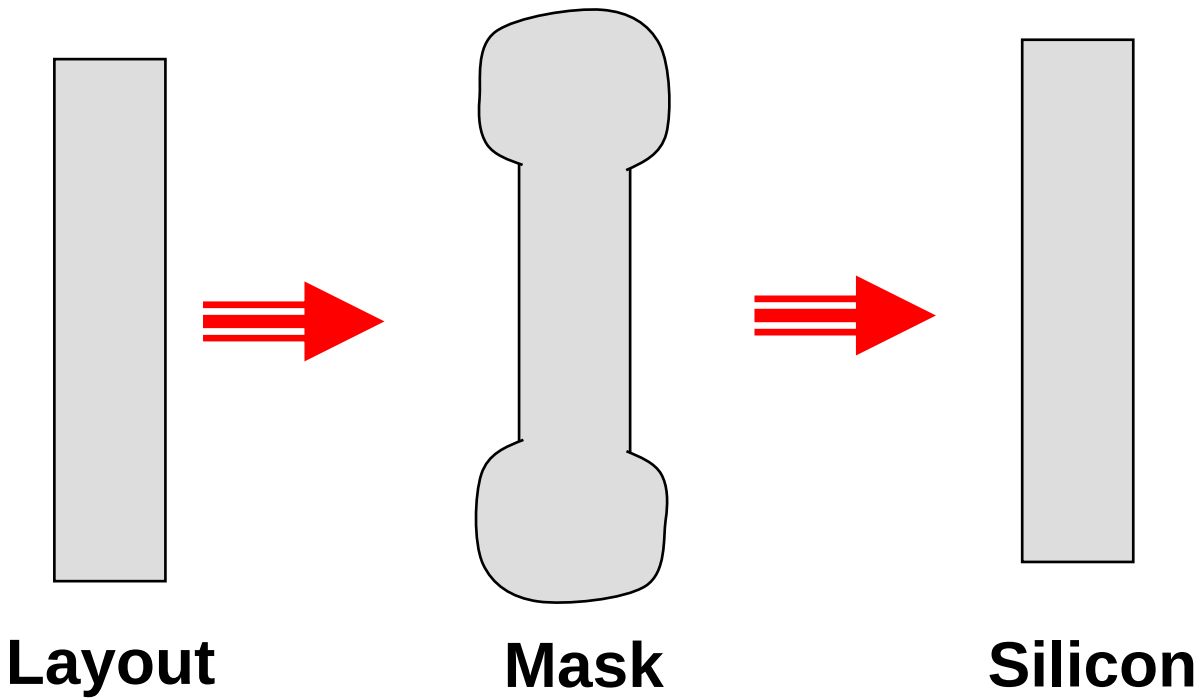
Accounting for VDSM Effects: Example

- **Example of 2D OPC effects: rounded edges, narrowed lines**



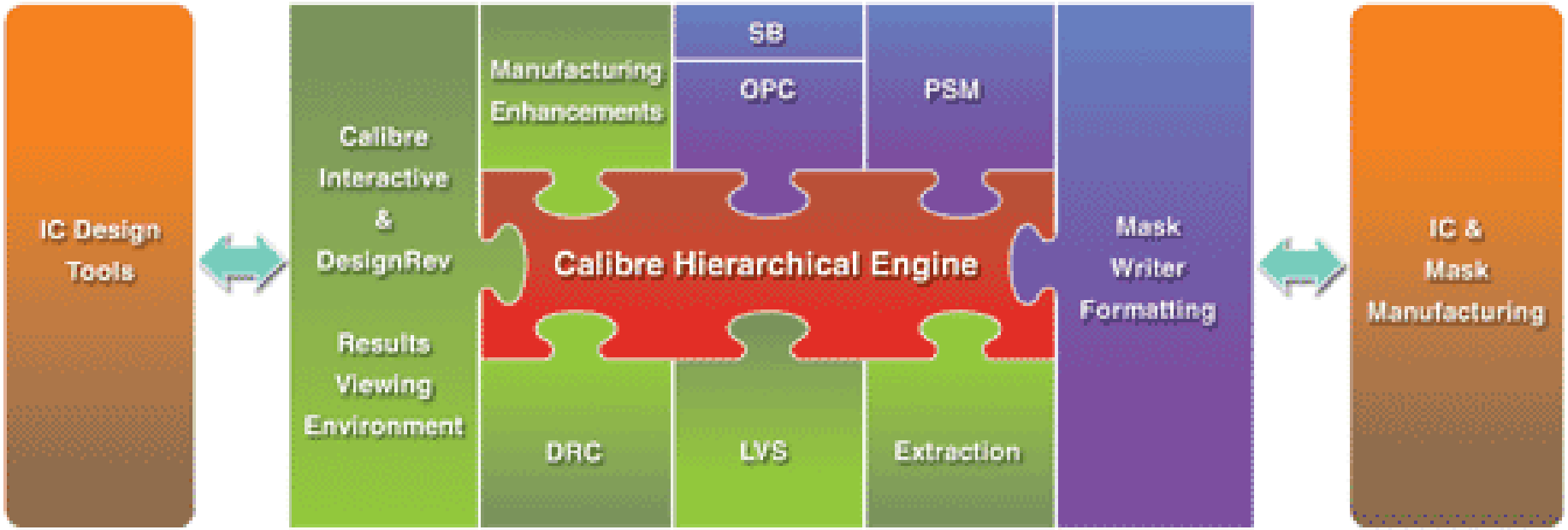
Uncorrected **Corrected**

Compensating for VDSM Effects: Masks



Compensating for VDSM Effects: CAD

- Flow to compensate is transparent to layout designer
- Layout design proceeds as normal



Mentor Graphics Flow

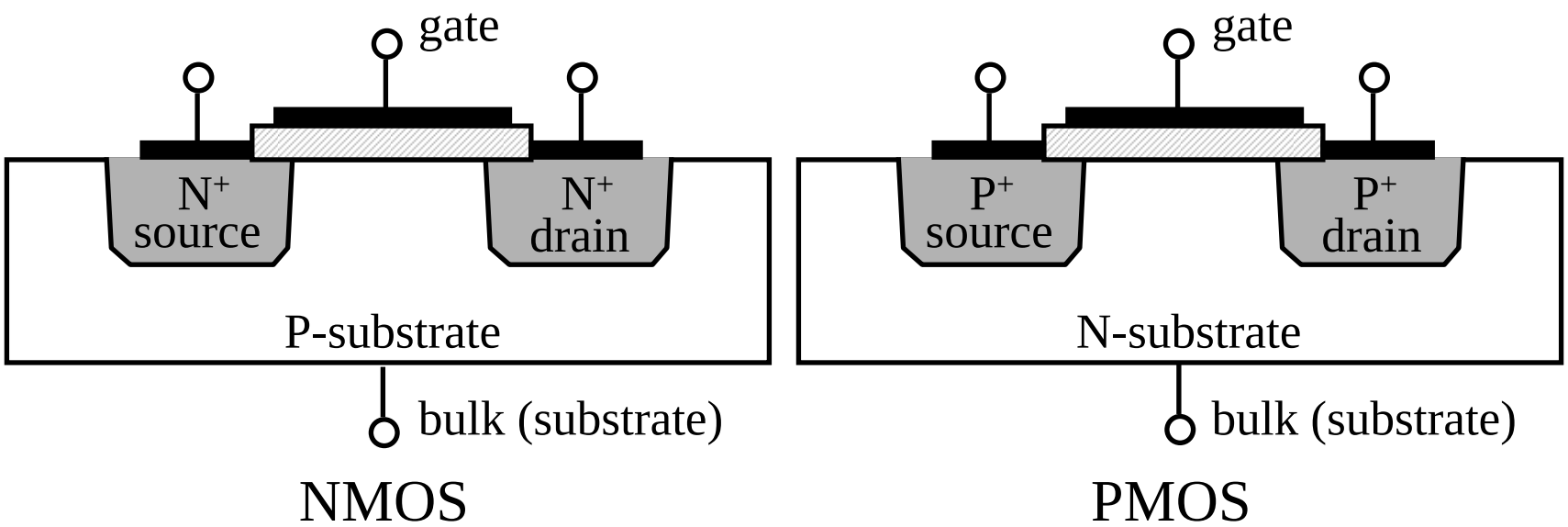
http://www.mentor.com/calibre/datasheets/opc/html/

References

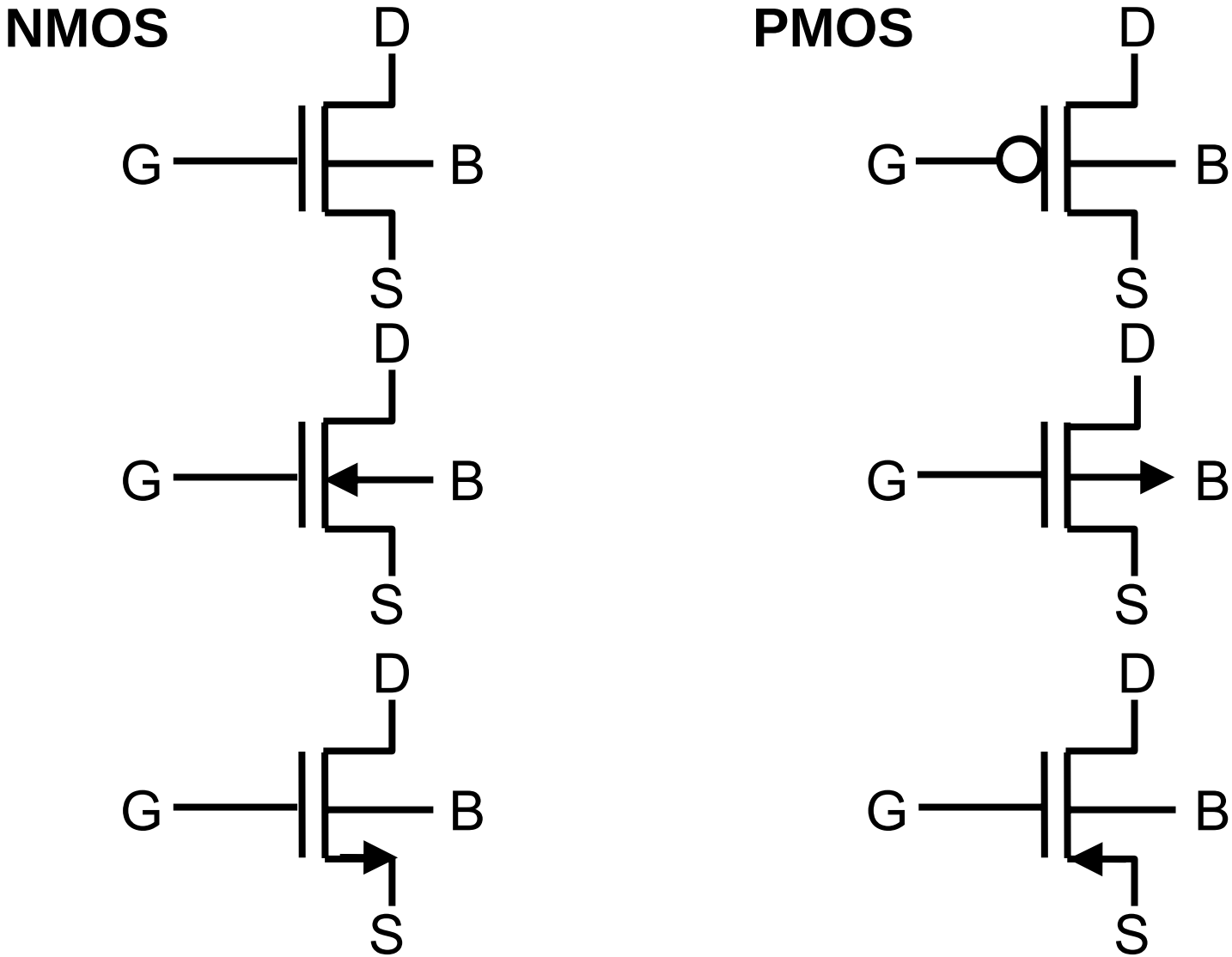
- **“Design of VLSI Systems”. A web based course located at: http://turquoise.wpi.edu/webcourse/**
- **“Simplified Rule Generation for Automated Rules-Based Optical Enhancement”, Otto et. al. On web at: http://www.jetlink.net/~ootto/bacus95/BACUS95Index.html**
- **Mark Anders and Jim Schantz of Intel Corporation**
- **Jan Rabaey, Lecture notes from his book “Digital Integrated Circuits, A Design Perspective”**

MOS Transistor Types

- **Rabaey Ch. 3 (Kang & Leblebici Ch. 3)**
- **Two transistor types (analogous to bipolar NPN, PNP)**
  - NMOS: p-type substrate, n^+ source/drain, electrons are charge carriers
  - PMOS: n-type substrate, p^+ source/drain, holes are charge carriers

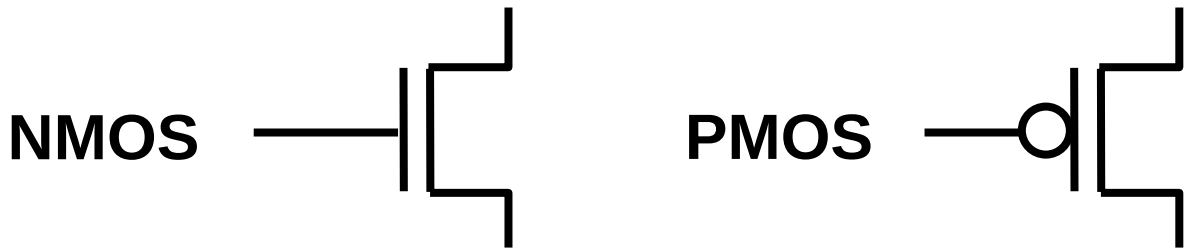


MOS Transistor Symbols



Note on MOS Transistor Symbols

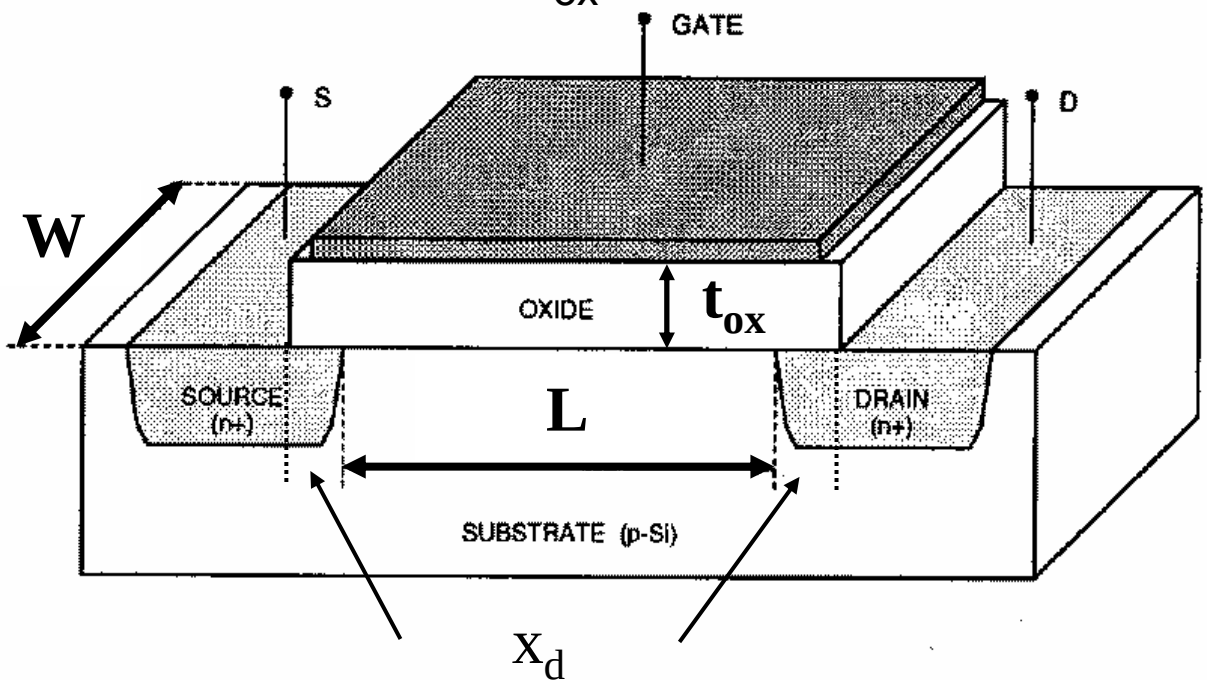
- **All symbols appear in literature**
  - Symbols with arrows are conventional in analog papers
  - PMOS with a bubble on the gate is conventional in digital circuits papers
- **Sometimes bulk terminal is ignored – implicitly connected to supply:**



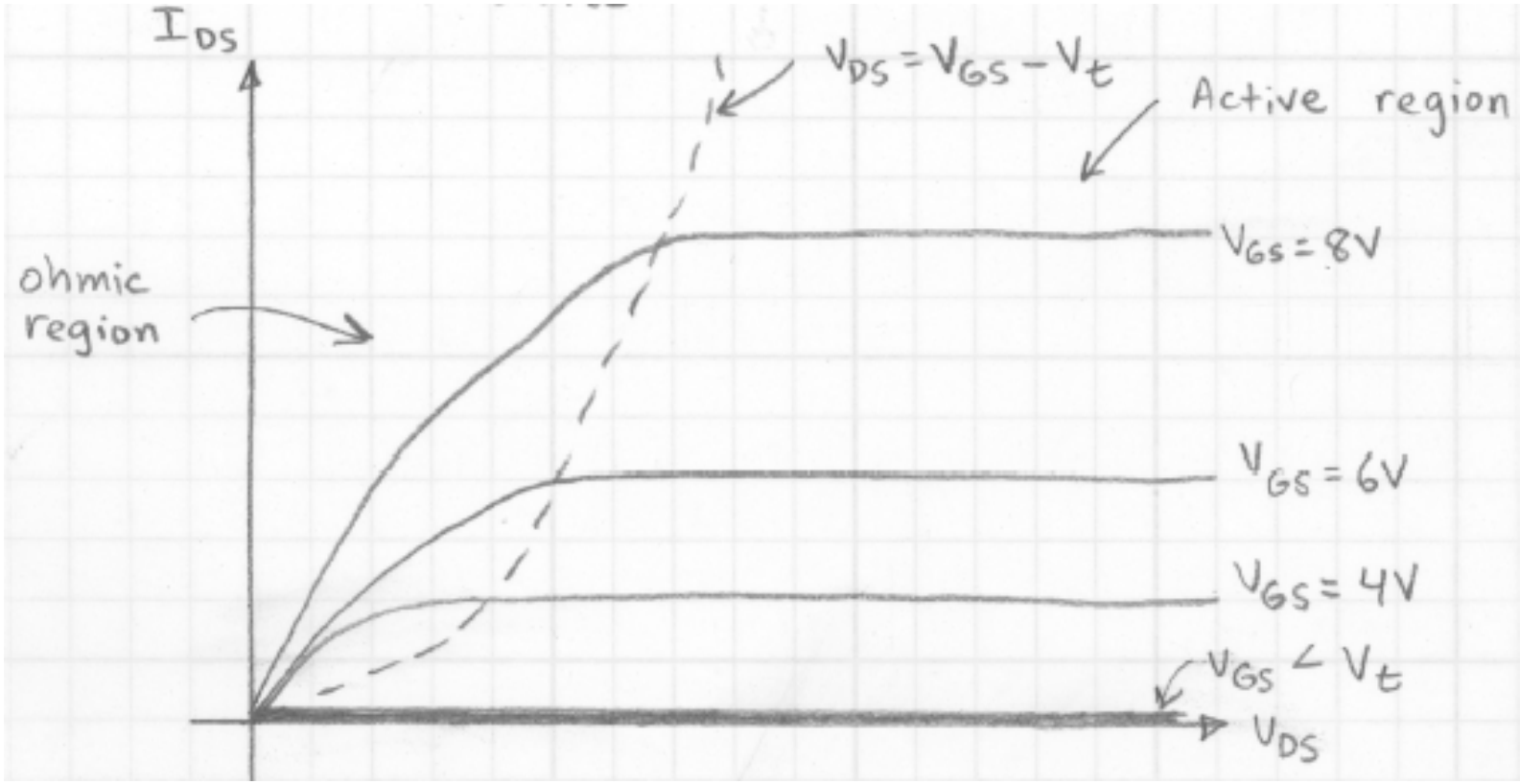
- **Unlike physical bipolar devices, source and drain are usually symmetric**

MOS Transistor Structure

- **Important transistor physical characteristics**
  - Channel length $L = L_D - 2x_d$ (K&L $L = Lgate - 2L_D$)
  - Channel width W
  - Thickness of oxide t_{ox}

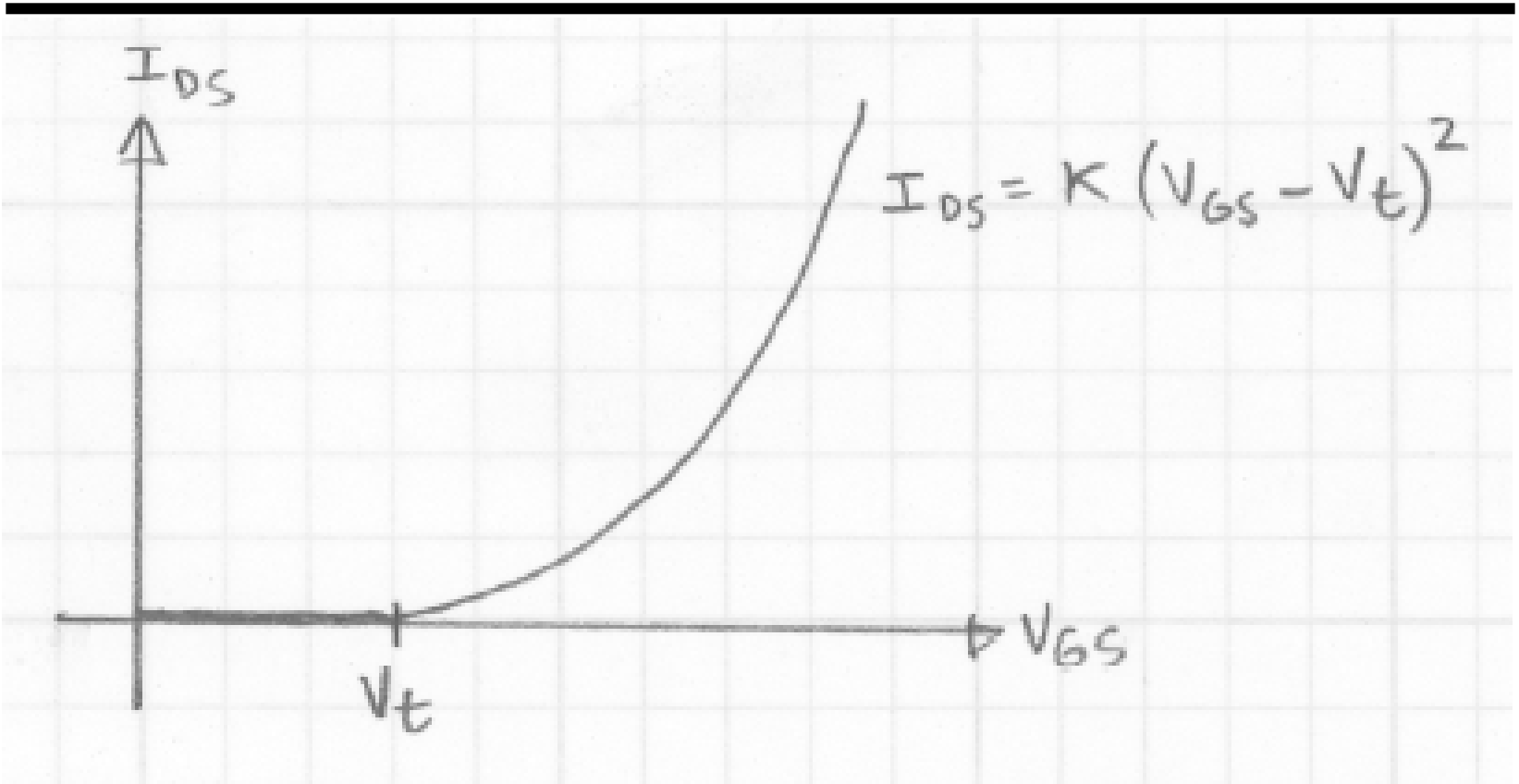


NMOS Transistor I-V Characteristics I



- **I-V curve vaguely resembles bipolar transistor curves**
  - Quantitatively very different
  - Turn-on voltage called Threshold Voltage V_T

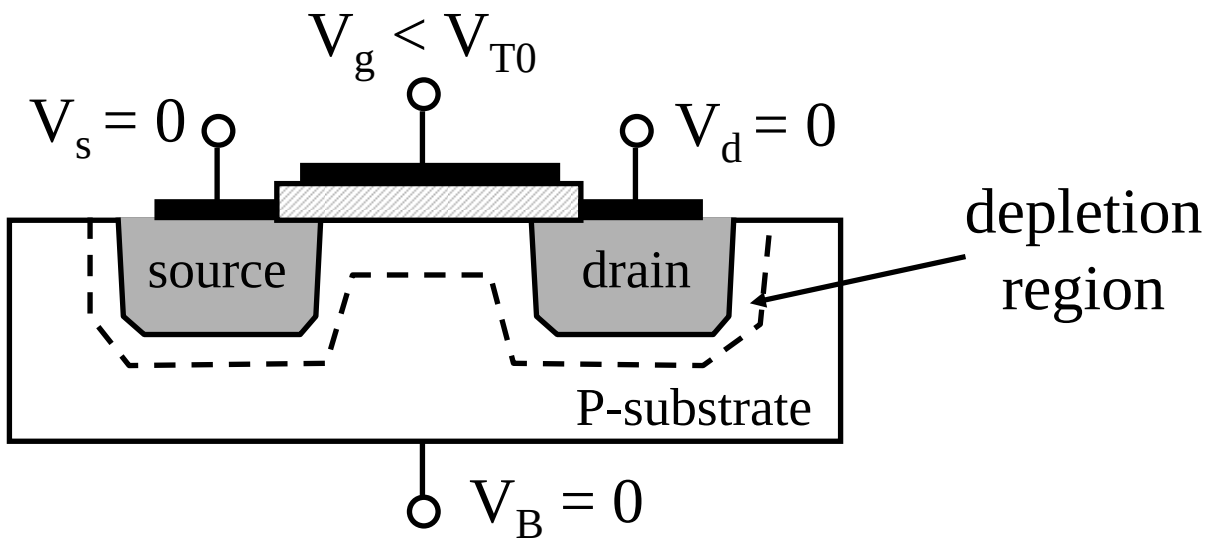
NMOS Transistor I-V Characteristics II



- **Drain current varies quadratically with gate-source voltage V_{GS} (in Saturation)**

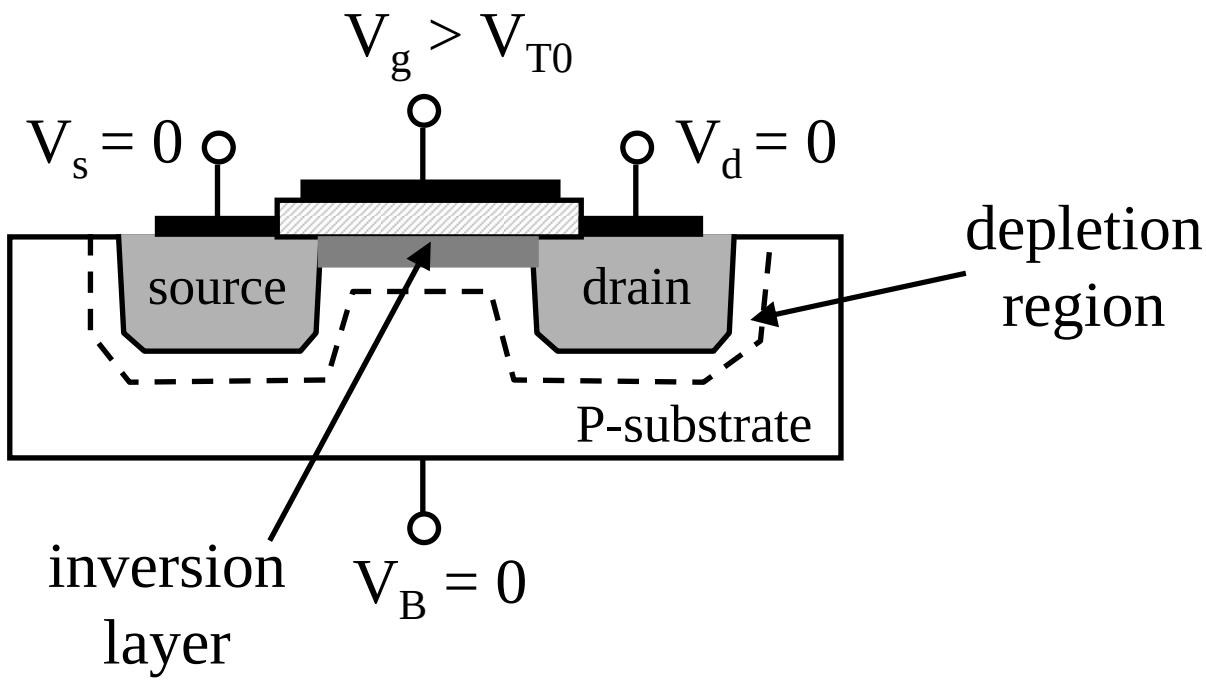
MOS Transistor Operation: Cutoff

- **Simple case: $V_D = V_S = V_B = 0$**
  - Operates as MOS capacitor (Cg = gate to channel)
  - Transistor in cutoff region
- **When $V_{GS} < V_{T0}$, depletion region forms**
  - No carriers in channel to connect S and D (Cutoff)



MOS Transistor Operation: Inversion

- **When $V_{GS} > V_{T0}$, inversion layer forms**
- **Source and drain connected by conducting n-type layer (for NMOS)**
  - Conducting p-type layer in PMOS



Threshold Voltage Components

- **Four physical components of the threshold voltage**

1. **Work function difference between gate and channel (depends on metal or polysilicon gate): Φ_{GC}**
2. **Gate voltage to invert surface potential: $-2\Phi_F$**
3. **Gate voltage to offset depletion region charge: Q_B/C_{ox}**
4. **Gate voltage to offset fixed charges in the gate oxide and oxide-channel interface: Q_{ox}/C_{ox}**

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$$ **: gate oxide capacitance per unit area**

Threshold Voltage Summary

- **If $V_{SB} = 0$ (no substrate bias):**

$$V_{T0} = \Phi_{GC} - 2\phi_F - \frac{Q_{B0}}{C_{ox}} - \frac{Q_{ox}}{C_{ox}} \quad \text{(K\&L 3.20)}$$

- **If $V_{SB} \neq 0$ (non-zero substrate bias)**

$$V_T = V_{T0} + \gamma\left(\sqrt{\left|-2\phi_F + V_{SB}\right|} - \sqrt{\left|2\phi_F\right|}\right) \quad (3.19)$$

- **Body effect (substrate-bias) coefficient:**

$$\gamma = \frac{\sqrt{2qN_A\varepsilon_{Si}}}{C_{ox}} \quad \text{(K\&L 3.24)}$$

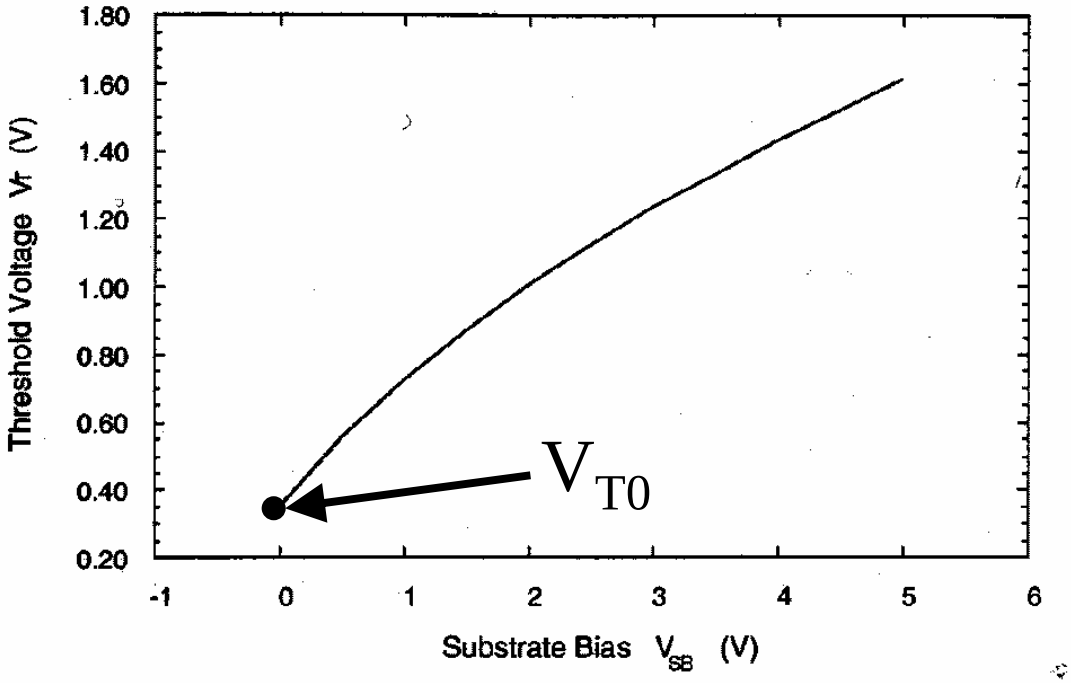
- **Threshold voltage increases as V_{SB} increases!**

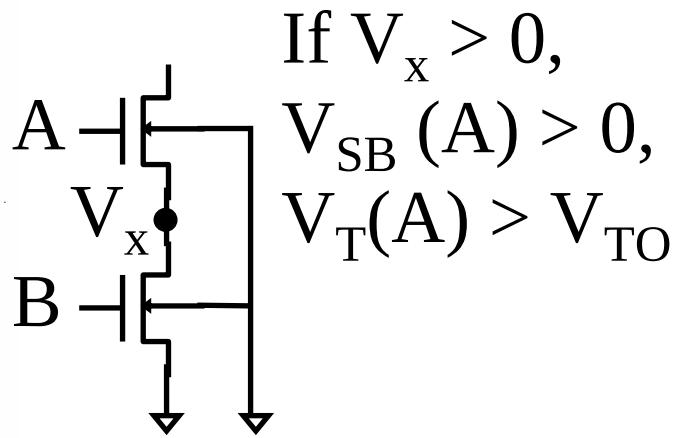
Threshold Voltage (NMOS vs. PMOS)

	NMOS	PMOS
Substrate Fermi potential	$\phi_F < 0$	$\phi_F > 0$
Depletion charge density	$Q_B < 0$	$Q_B > 0$
Substrate bias coefficient	$\gamma > 0$	$\gamma < 0$
Substrate bias voltage	$V_{SB} > 0$	$V_{SB} < 0$

Body Effect

- **Body effect: Source-bulk voltage V_{SB} affects threshold voltage of transistor**
  - Body normally connected to ground for NMOS, Vdd (Vcc) for PMOS
  - Raising source voltage increases V_T of transistor
  - Implications on circuit design: series stacks of devices





MOS Transistor Regions of Operation

- **Three main regions of operation**
- **<u>Cutoff</u>: $V_{GS} < V_T$**
  No inversion layer formed, drain and source are isolated by depleted channel. $I_{DS} \approx 0$
- **<u>Linear (Triode, Ohmic)</u>: $V_{GS} > V_T$, $V_{DS} < V_{GS}-V_T$**
  Inversion layer connects drain and source. Current is almost linear with V_{DS} (like a resistor)
- **<u>Saturation</u>: $V_{GS} > V_T$, $V_{DS} \geq V_{GS}-V_T$**
  Channel is "pinched-off". Current saturates (becomes independent of V_{DS}, to first order).

MOSFET Drain Current Overview

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

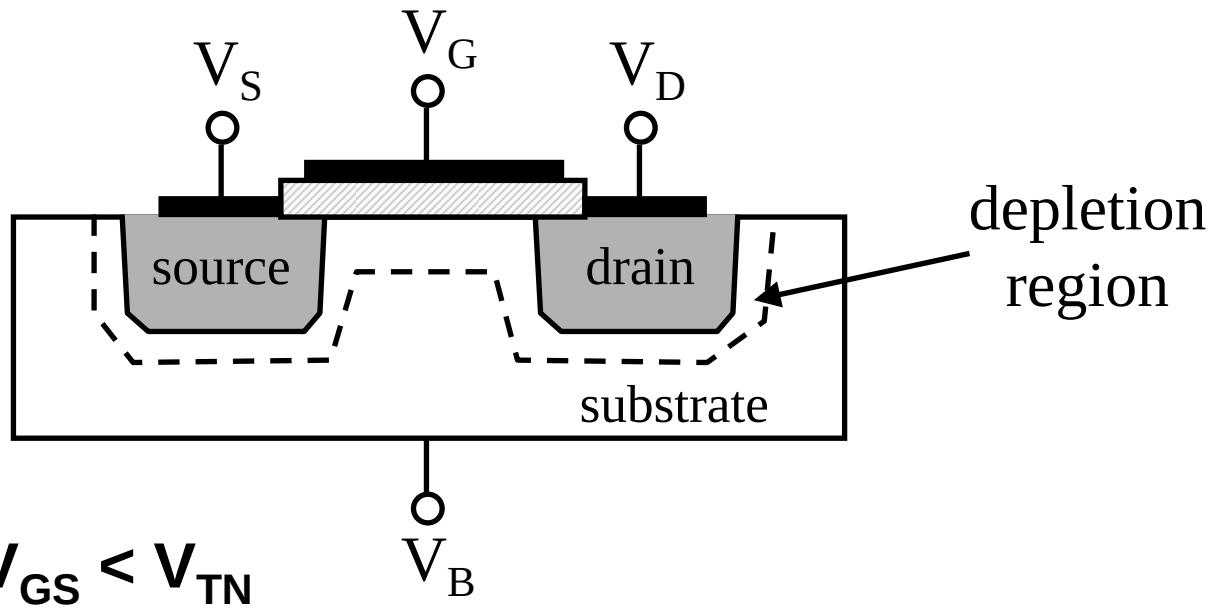
Linear (Triode, Ohmic):

$$I_D = \mu C_{ox}\frac{W}{L}\left(\left(V_{GS} - V_T\right)V_{DS} - \frac{V_{DS}^2}{2}\right)$$

Cutoff: $$I_D \approx 0$$

"Classical" MOSFET model, will discuss deep submicron modifications as necessary (Rabaey, Eqs. 3.25, 3.29)

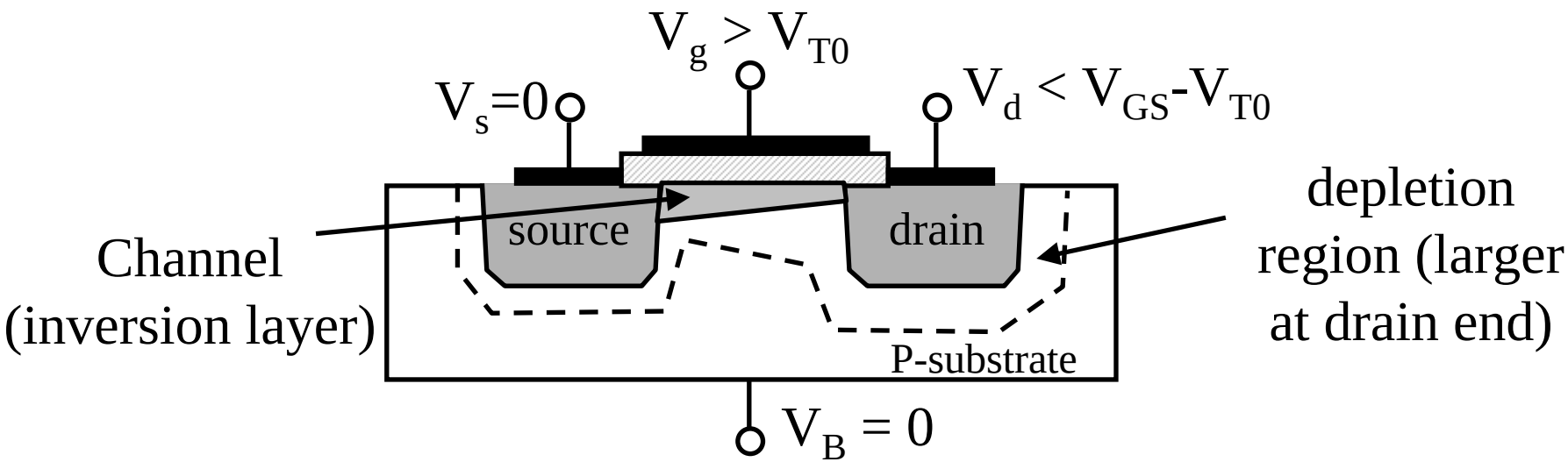
Cutoff Region



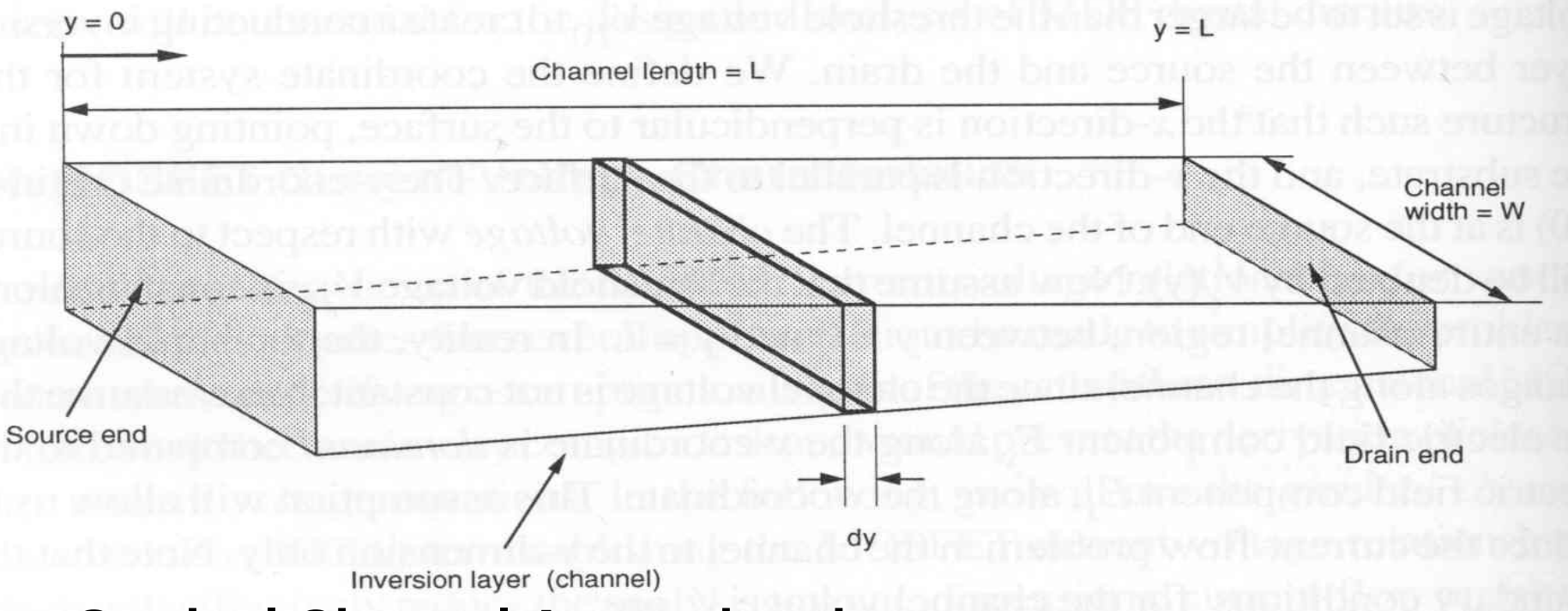
- **For NMOS: $V_{GS} < V_{TN}$**
- **For PMOS: $V_{GS} > V_{TP}$**
- **Depletion region – no inversion**
- **Current between drain and source is 0**
  - Actually there is always some leakage (subthreshold) current

Linear Region

- **When $V_{GS}>V_T$, an inversion layer forms between drain and source**
- **Current I_{DS} flows from drain to source (electrons travel from source to drain)**
- **Depth of channel depends on V between gate and channel**
  - Drain end narrower due to larger drain voltage
  - Drain end depth reduces as V_{DS} is increased



Linear Region I/V Equation Derivation



- **Gradual Channel Approximation:**
  - Assume dominant electric field in y-direction
  - Current is constant along channel
- **Integrate differential voltage drop $dV_c = I_D dR$ along y**

Linear Region I/V Equation

- **Valid for continuous channel from Source to Drain**

$$I_D = \mu_n C_{ox} \frac{W}{L}\left[(V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2\right]$$

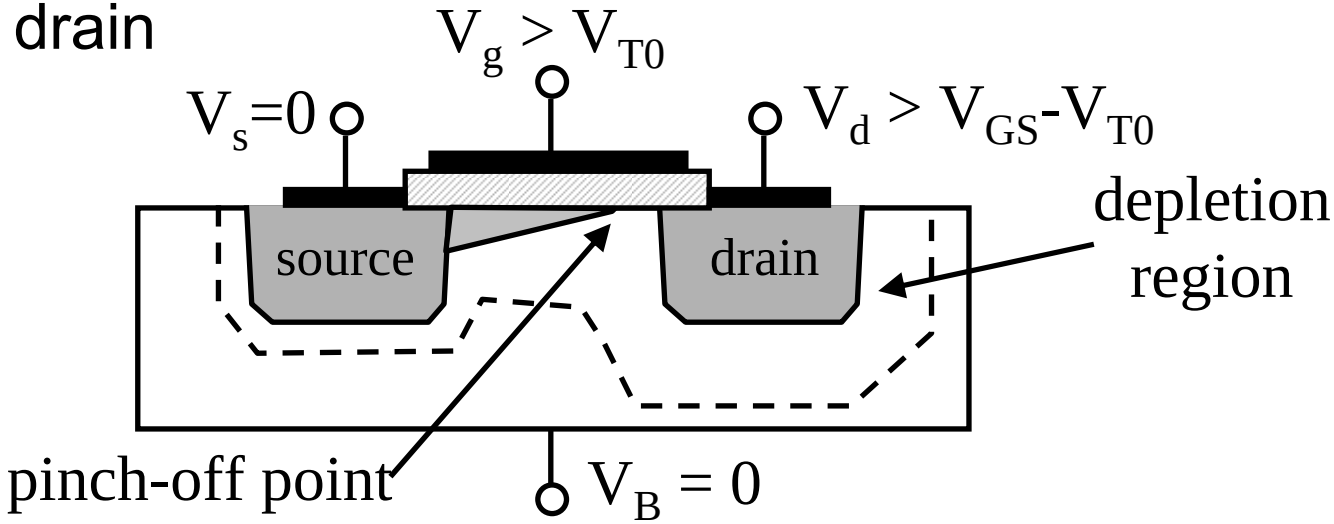
Device transconductance: $k_n = \mu_n C_{ox} \frac{W}{L}$

Process transconductance: $k_n' = \mu_n C_{ox}$

$$I_D = k_n' \frac{W}{L}\left[(V_{GS} - V_T)V_{DS} - \tfrac{1}{2}V_{DS}^2\right]$$

Saturation Region

- **When $V_{DS} = V_{GS} - V_T$:**
  - No longer voltage drop of V_T from gate to substrate at drain
  - Channel is "pinched off"
- **If V_{DS} is further increased, no increase in current I_{DS}**
  - As V_{DS} increased, pinch-off point moves closer to source
  - Channel between that point and drain is depleted
  - High electric field in depleted region accelerates electrons towards drain



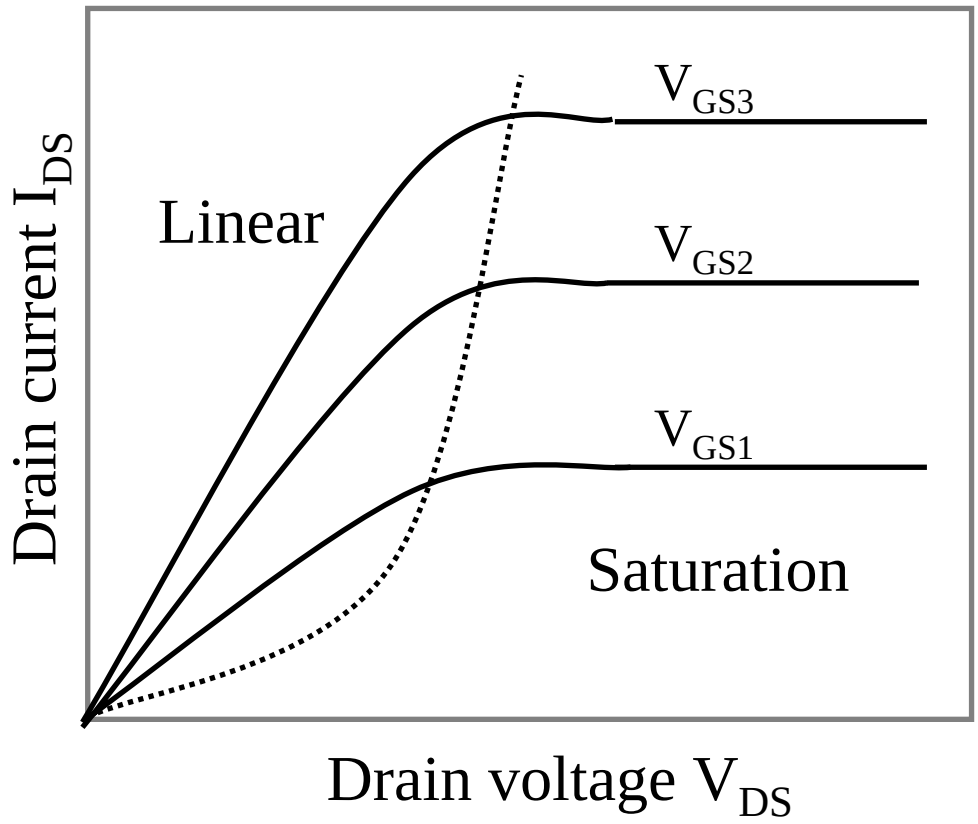
Saturation I/V Equation

- **As drain voltage increases, channel remains pinched off**
  - Channel voltage remains constant
  - Current saturates (no increase with increasing V_{DS})
- **To get saturation current, use linear equation with $V_{DS} = V_{GS} - V_T$**

$$I_D = \frac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{GS} - V_{TN}\right)^2$$

MOS I/V Characteristics

- **I/V curve for ideal MOS device**
- **$V_{GS3} > V_{GS2} > V_{GS1}$**



Channel Length Modulation

- **In saturation, pinch-off point moves**
  - As V_{DS} is increased, pinch-off point moves closer to source
  - Effective channel length becomes shorter
  - Current increases due to shorter channel

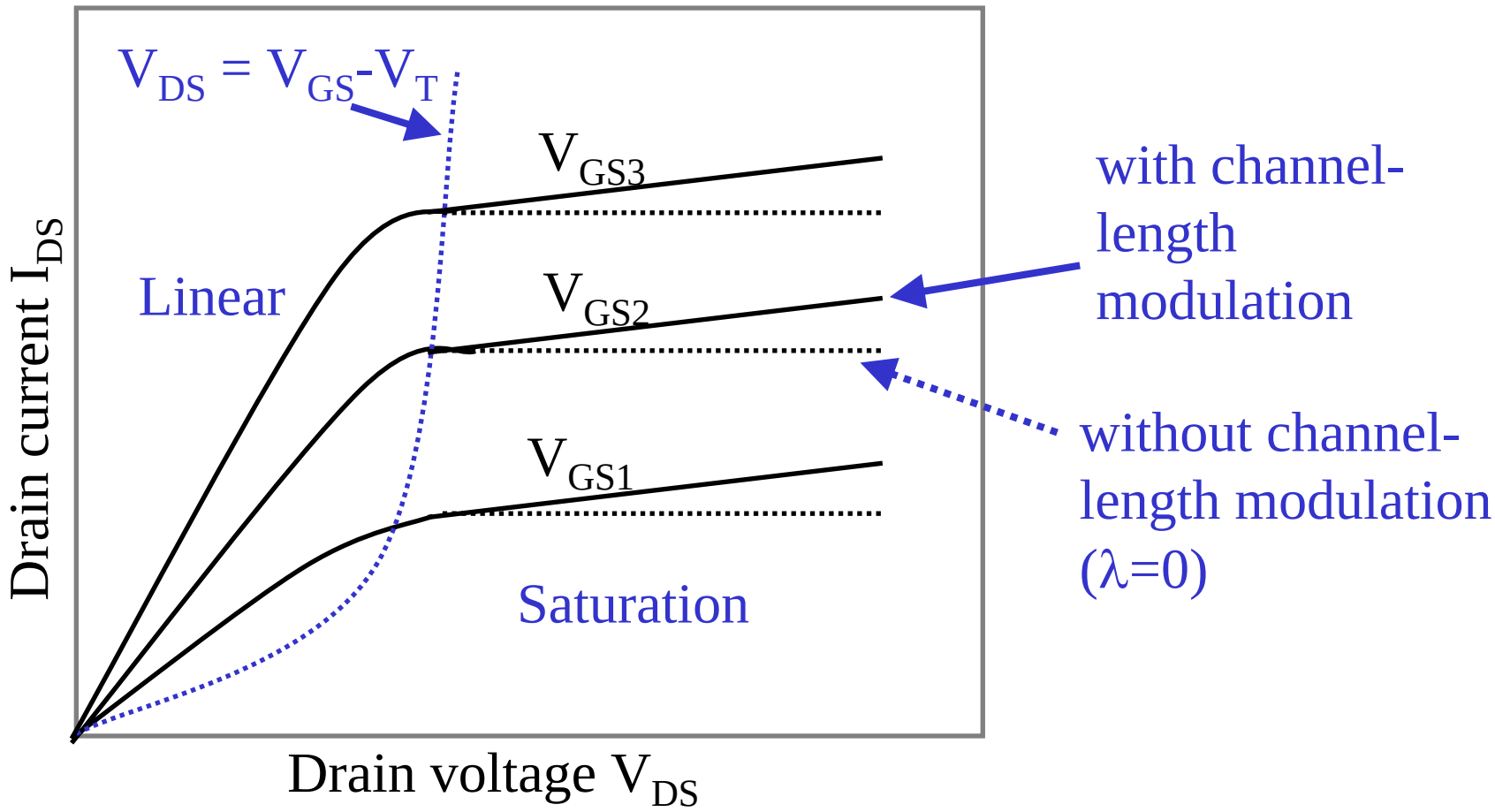
$$L' = L - \Delta L$$

$$I_D = \tfrac{1}{2}\mu_n C_{ox}\frac{W}{L}\left(V_{GS} - V_{TN}\right)^2\left(1 + \lambda V_{DS}\right)$$

λ = channel length modulation coefficient

MOS I/V Curve Summary

I/V curve for non-ideal NMOS device:



MOS I/V Equations Summary

Cutoff

$$\begin{aligned} V_{GS} &< V_{TN} \\ V_{GS} &> V_{TP} \end{aligned} \Rightarrow I_D = 0$$

Linear

$$\begin{aligned} V_{GS} \geq V_{TN}, &\quad V_{DS} < V_{GS} - V_{TN} \\ V_{GS} \leq V_{TP}, &\quad V_{DS} > V_{GS} - V_{TP} \end{aligned} \Rightarrow I_D = \mu C_{ox} \frac{W}{L} \left[\left(V_{GS} - V_T \right) V_{DS} - \tfrac{1}{2} V_{DS}^2 \right]$$

Saturation

$$\begin{aligned} V_{GS} \geq V_{TN}, &\quad V_{DS} \geq V_{GS} - V_{TN} \\ V_{GS} \leq V_{TP}, &\quad V_{DS} \leq V_{GS} - V_{TP} \end{aligned} \Rightarrow I_D = \tfrac{1}{2} \mu C_{ox} \frac{W}{L} \left(V_{GS} - V_T \right)^2 \left(1 + \lambda V_{DS} \right)$$

Note: if $V_{SB} \neq 0$, need to recalculate V_T from V_{T0}

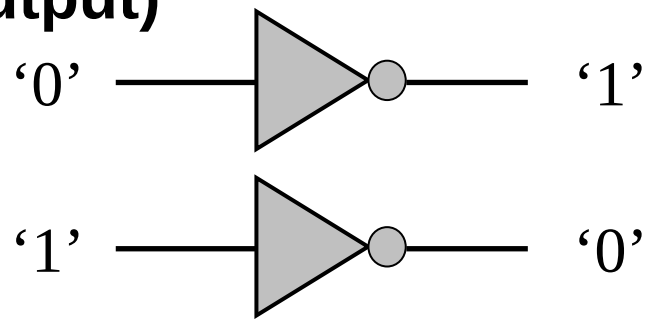
A Fourth Region: Subthreshold

Subthreshold: $$I_D = I_S e^{\frac{V_{GS}}{n kT/q}} \left(1 - e^{-\frac{V_{DS}}{kT/q}} \right)$$

- **Sometimes called "weak inversion" region**
- **When V_{GS} near V_T, drain current has an exponential dependence on gate to source voltage**
  - Similar to a bipolar device
- **Not typically used in digital circuits**
  - Sometimes used in very low power digital applications
  - Often used in low power analog circuits, e.g. quartz watches

Inverter Operation

- **Inverter is simplest digital logic gate (1 input, 1 output)**

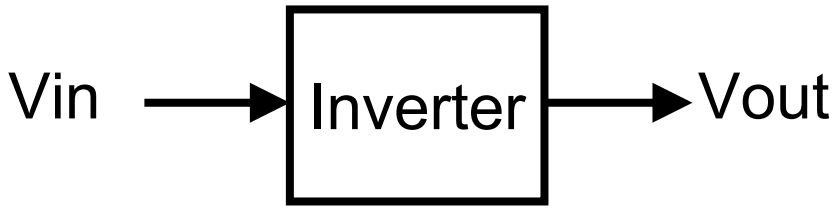


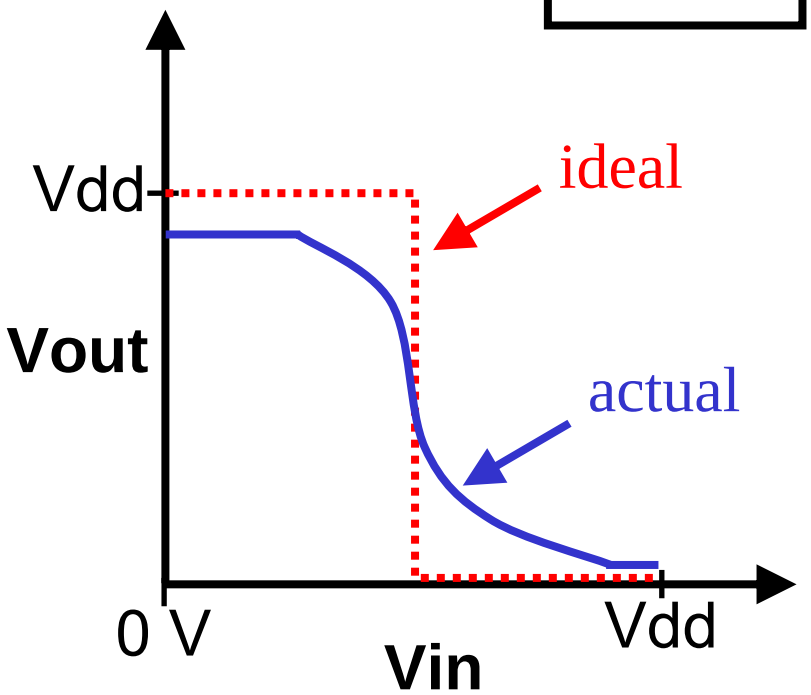
In	Out
0	1
1	0

- **Many different circuit styles possible**
  - Resistive-load
  - NMOS and Pseudo-NMOS
  - CMOS
- **Important static and dynamic characteristics**
  - Speed (delay through the gate)
  - Power consumption
  - Robustness (tolerance to noise)
  - Area and process cost

Inverter Model: Voltage Transfer Curve

Voltage transfer curve (VTC): plot of output voltage Vout vs. input voltage Vin

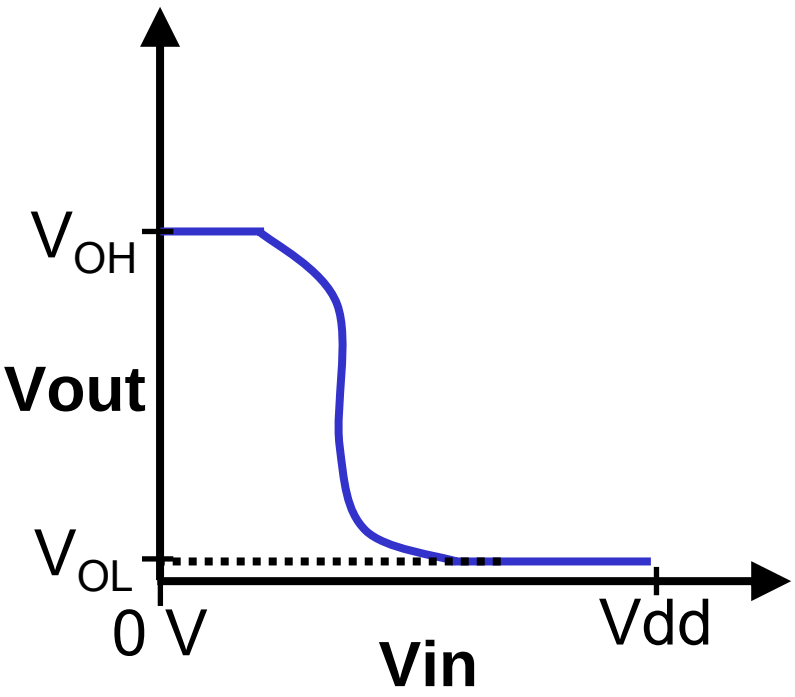




Ideal digital inverter:

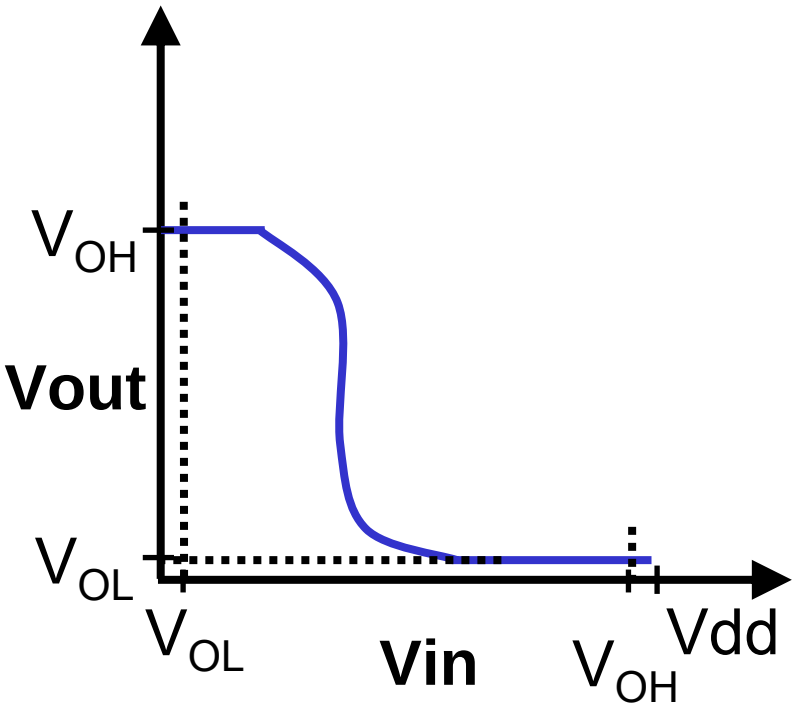
- When Vin=0, Vout=Vdd
- When Vin=Vdd, Vout=0
- Infinitely sharp transition region at inverter switching threshold

Actual Inverter: V_{OH} and V_{OL}



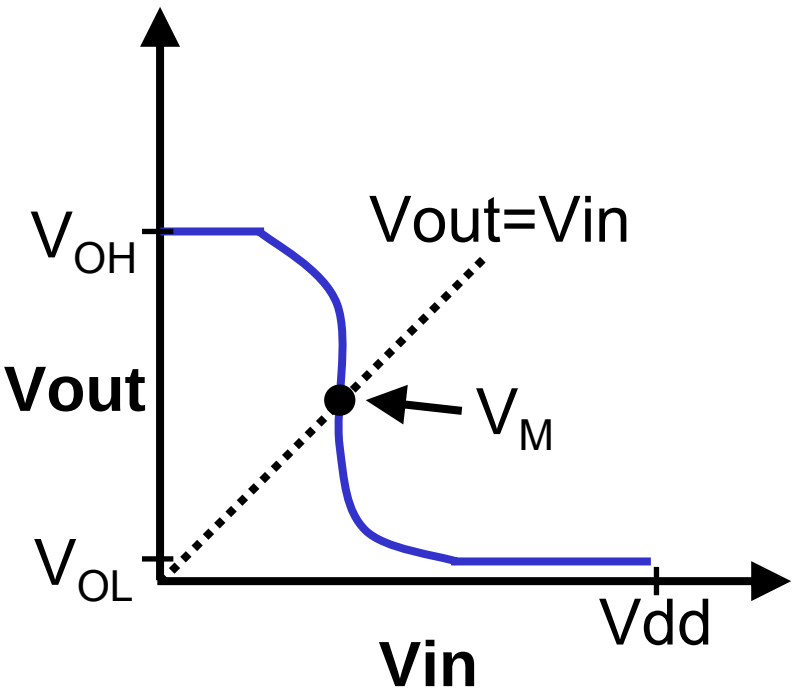
- **V_{OH} and V_{OL} represent the "high" and "low" output voltages of the inverter**
- **V_{OH} = output voltage when Vin = '0' (V Output High)**
- **V_{OL} = output voltage when Vin = '1' (V Output Low)**
- **Ideally,**
  - V_{OH} = Vdd
  - V_{OL} = 0 V

VOL and VOH



- **In transfer function terms:**
  - $V_{OL} = f(V_{OH})$
  - $V_{OH} = f(V_{OL})$
  - f = inverter transfer function
- **Difference (V_{OH}-V_{OL}) is the *voltage swing* of the gate**
  - *Full-swing logic* swings from ground to Vdd
  - Other families with smaller swings

Inverter Switching Threshold

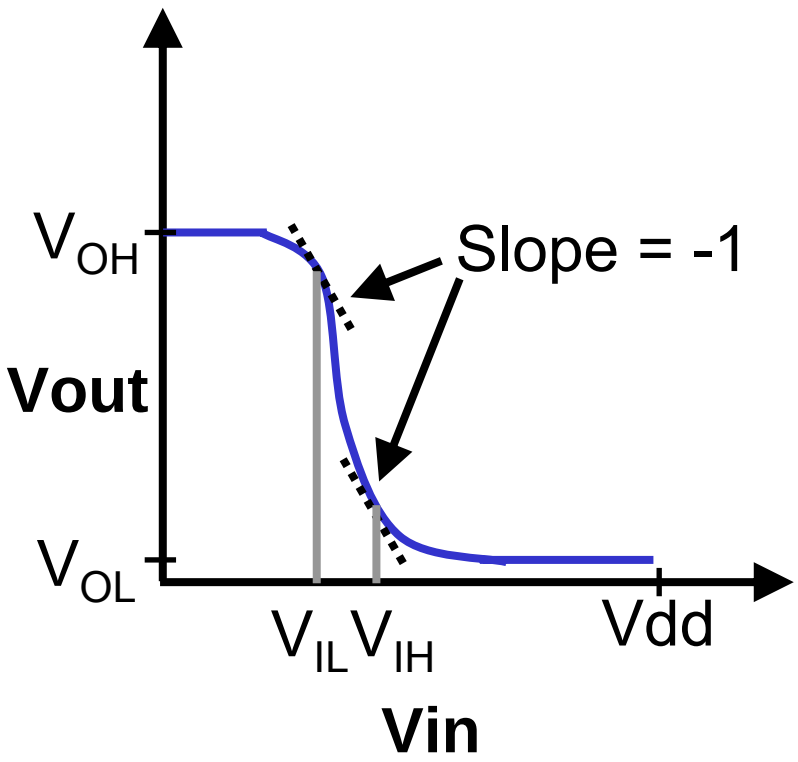


Inverter switching threshold:

- Point where voltage transfer curve intersects line Vout=Vin
- Represents the point at which the inverter switches state
- Normally, $V_M \approx Vdd/2$
- Sometimes other thresholds desirable

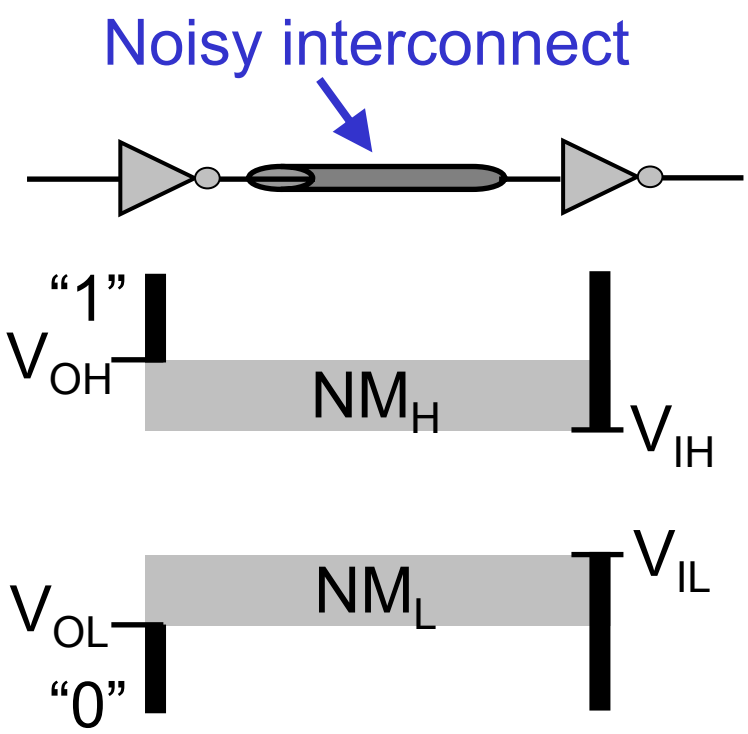
(K&L $V_{TH} = V_M$)

Noise Margins



- **V_{IL} and V_{IH} measure effect of input voltage on inverter output**
- **V_{IL} = largest input voltage recognized as logic ‘0’**
- **V_{IH} = smallest input voltage recognized as logic ‘1’**
- **Defined as point on VTC where slope = -1**

Noise Margins and Robustness



Ideally, noise margin should be as large as possible

- **Noise margin is a measure of the *robustness* of an inverter**
  - $N_{ML} = V_{IL} - V_{OL}$
  - $N_{MH} = V_{OH} - V_{IH}$
- **Models a chain of inverters. Example:**
  - First inverter output is V_{OH}
  - Second inverter recognizes input $> V_{IH}$ as logic '1'
  - Difference V_{OH}-V_{IH} is "safety zone" for noise

Noise Margin Motivation

- **Why are V_{IL}, V_{IH} defined as unity-gain point on VTC curve?**
  - Assume there is noise on input voltage V_{in}

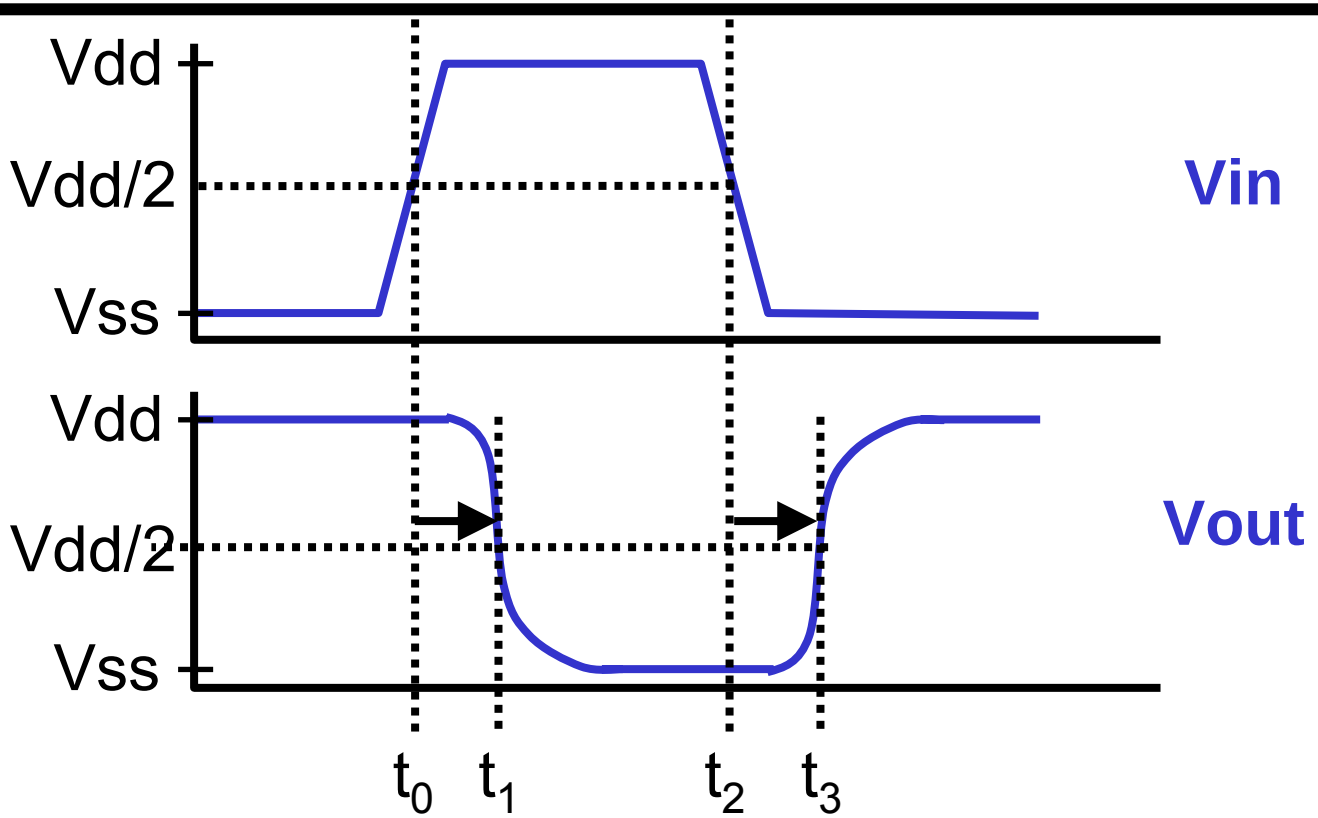
$$V_{out} = f\left(V_{in} + V_{noise}\right)$$

  - First-order Taylor series approximation:

$$V_{out} = f\left(V_{in}\right) + \frac{dV_{out}}{dV_{in}} V_{noise}$$

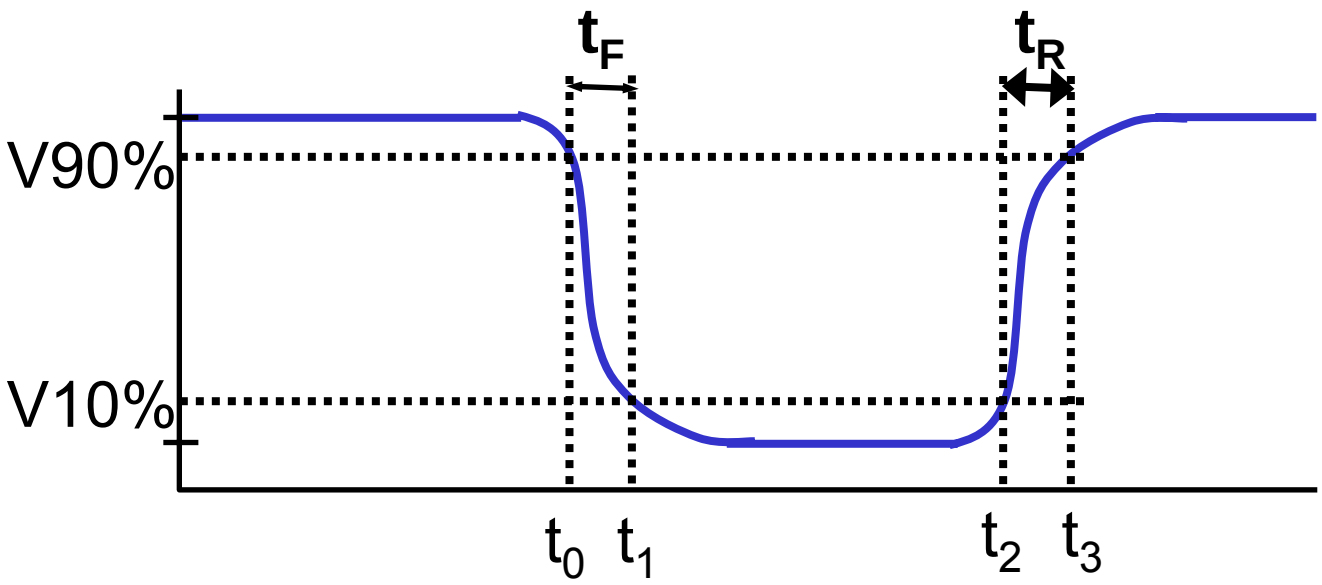
  - If gain (dV_{out}/dV_{in}) > 1, noise will be amplified.
  - If gain < 1, noise is filtered. Therefore V_{IL}, V_{IH} define regions where gain < 1

Inverter Time Response



- **Propagation delay measured from 50% point of Vin to 50% point of Vout**
- **$t_{phl} = t_1 - t_0$, $t_{plh} = t_3 - t_2$, $t_p = \frac{1}{2}(t_{phl}+t_{plh})$**

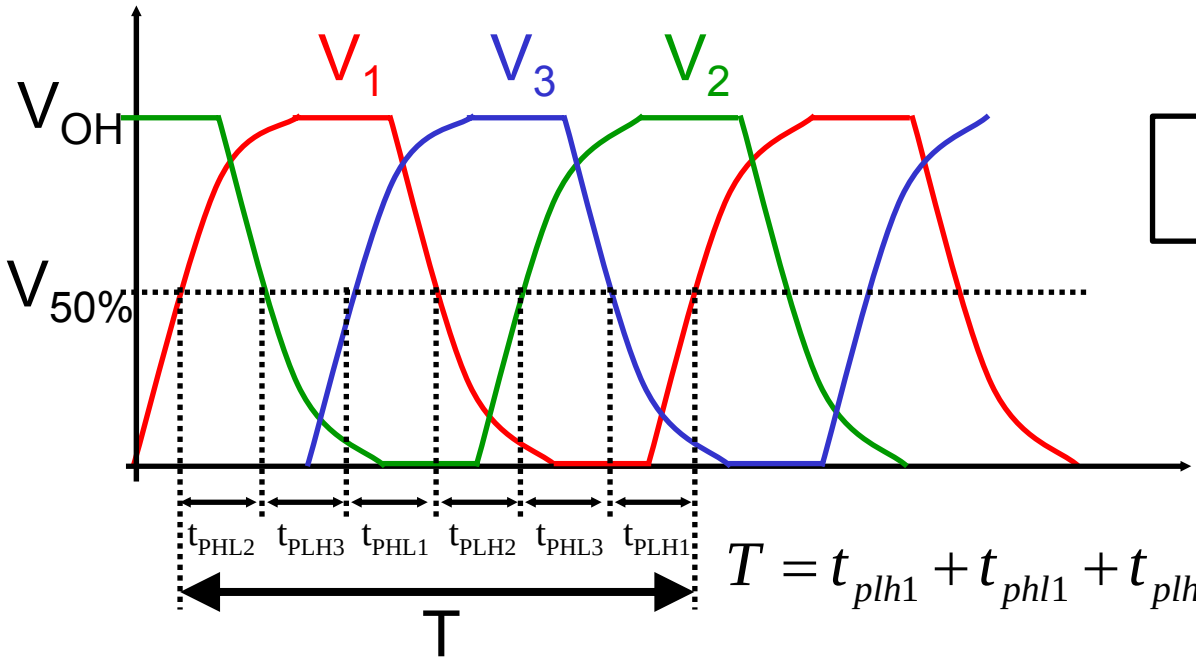
Rise and Fall Time

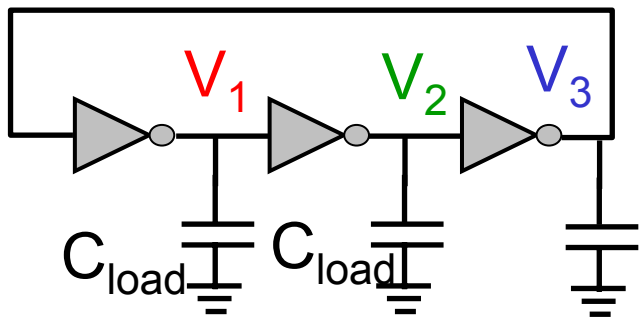


- **Fall time: measured from 90% point to 10% point**
  - $t_F = t_1 - t_0$
- **Rise time: measured from 10% point to 90% point**
  - $t_R = t_3 - t_2$
- **Alternately, can define 20%-80% rise/fall time**

Ring Oscillator

- ***Ring oscillator circuit*: standard method of comparing delay from one process to another**
- **Odd-number n of inverters connected in chain: oscillates with period T (usually n >> 5)**

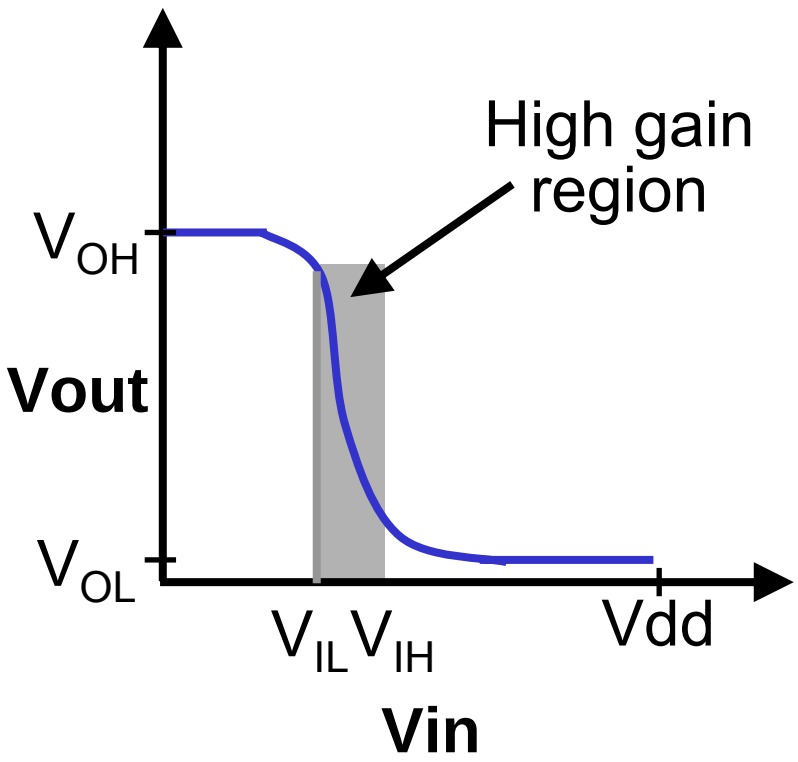




$$T = t_{plh1} + t_{phl1} + t_{plh2} + t_{phl2} + t_{plh3} + t_{phl3} + \cdots$$

$$T = 2nt_p, \qquad f = \frac{1}{T} = \frac{1}{2nt_p}, \qquad t_p = \frac{1}{2nf}$$

Inverter as Amplifier



- For V_{in} between V_{IL} and V_{IH}, inverter gain > 1
- Acts as a linear amplifier (often very high gain)
- Logic levels '0' and '1' correspond to saturating amplifier output (output is pegged to high or low supply)
- Resistive load inverter same circuit as common source amplifier

Next Topic: Inverters

- **Inverter Characteristics**
  - Transfer functions, noise margins, resistive and nonlinear loads
- **CMOS Inverters**
- **MOSFET Scaling**
- **MOSFET Capacitances**