BIRZEIT UNIVERSITY

# DEPARTMENT OF COMPUTER SYSTEM ENGINEERING

## Digital Integrated Circuits - ENCS333

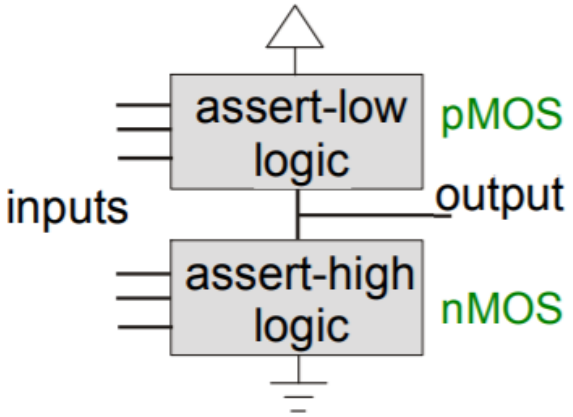
**Dr. Khader Mohammad**

**Lecture #6-** CMOS Inverter

## Integrated-Circuit Devices and Modeling

# Creating Logic Gates in CMOS

- All standard Boolean logic functions (INV, NAND, OR, etc.) can be produced in CMOS push-pull circuits.
- Rules for constructing logic gates using CMOS
  - use a complementary nMOS/pMOS pair for each input
  - connect the output to VDD through pMOS txs
  - connect the output to ground through nMOS txs
  - ensure the output is always either high or low
- CMOS produces “inverting” logic
  - CMOS gates are based on the inverter
  - outputs are always inverted logic functions e.g., NOR, NAND rather than OR, AND




## Logic Properties

***DeMorgan’s Rules***

$(a \cdot b)' = a' + b'$

$(a + b)' = a' \cdot b'$

***Useful Logic Properties***

$1 + x = 1 \quad 0 + x = x$

$1 \cdot x = x \quad 0 \cdot x = 0$

$x + x' = 1 \quad x \cdot x' = 0$

$a \cdot a = a \quad a + a = a$

$ab + ac = a\,(b+c)$

***Properties which can be proven***

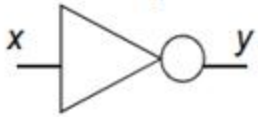
$(a+b)(a+c) = a+bc$

$a + a'b = a + b$

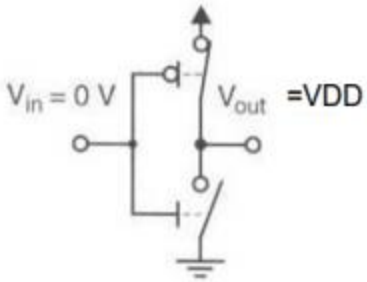
# CMOS Inverter

- Inverter Function
  - toggle binary logic of a signal

- Inverter Symbol



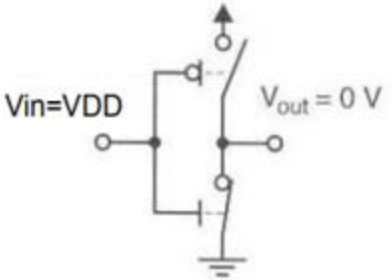

- Inverter Switch Operation




(a)

input low → output high
nMOS off/open
pMOS on/closed

pMOS "on"
→ output high (1)




(b)

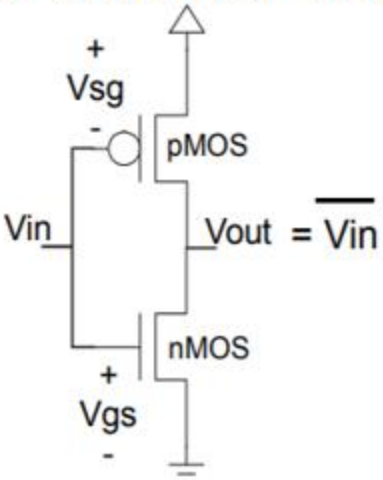
input high → output low
nMOS on/closed
pMOS off/open

nMOS "on"
→ output low (0)

- Inverter Truth Table

| $x$ | $y = \overline{x}$ |
|---|---|
| 0 | 1 |
| 1 | 0 |

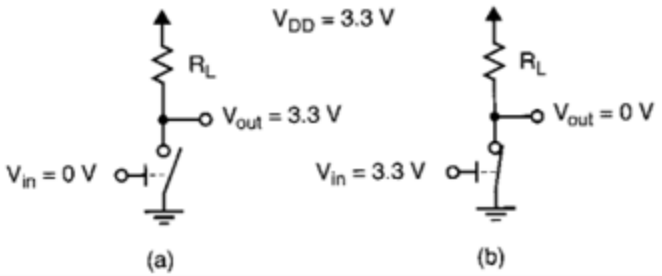
- CMOS Inverter Schematic



# nMOS Logic Gates

- Study nMOS logic first, more simple than CMOS
- nMOS Logic
  - assume a resistive load to VDD
  - nMOS switches pull output low based on inputs

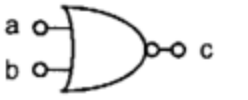
nMOS Inverter



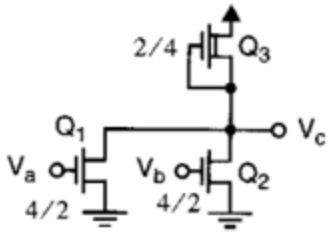

(a) nMOS is off
→ output is high (1)

(b) nMOS is on
→ output is low (0)

nMOS NOR



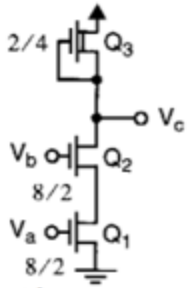

$c = \overline{a+b}$




- parallel switches = OR function
- nMOS pulls low (NOTs the output)

nMOS NAND


$c = \overline{ab}$




- series switches = AND function
- nMOS pulls low (NOTs the output)

# Abstraction Level

High

Structural Verilog:

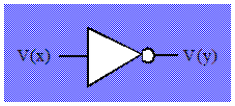
```
wire s;
wire sbar;
NOT not1(s, sbar);
```

Architecture SPEC

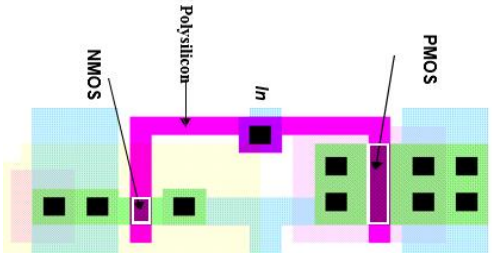
Behavioral Verilog:

```
wire s, sbar;
sbar = ~s;
```

RTL




Schematics

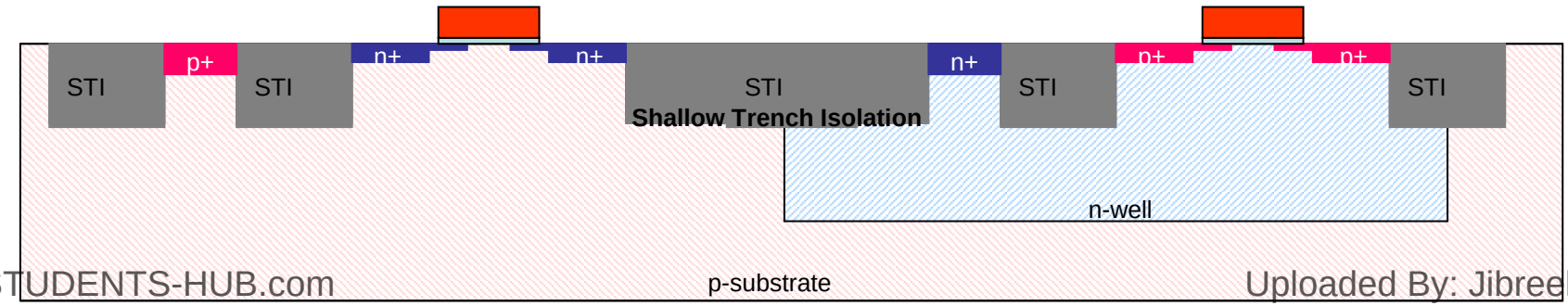



Physical Design (Layout)

Silicon

Abstraction Level

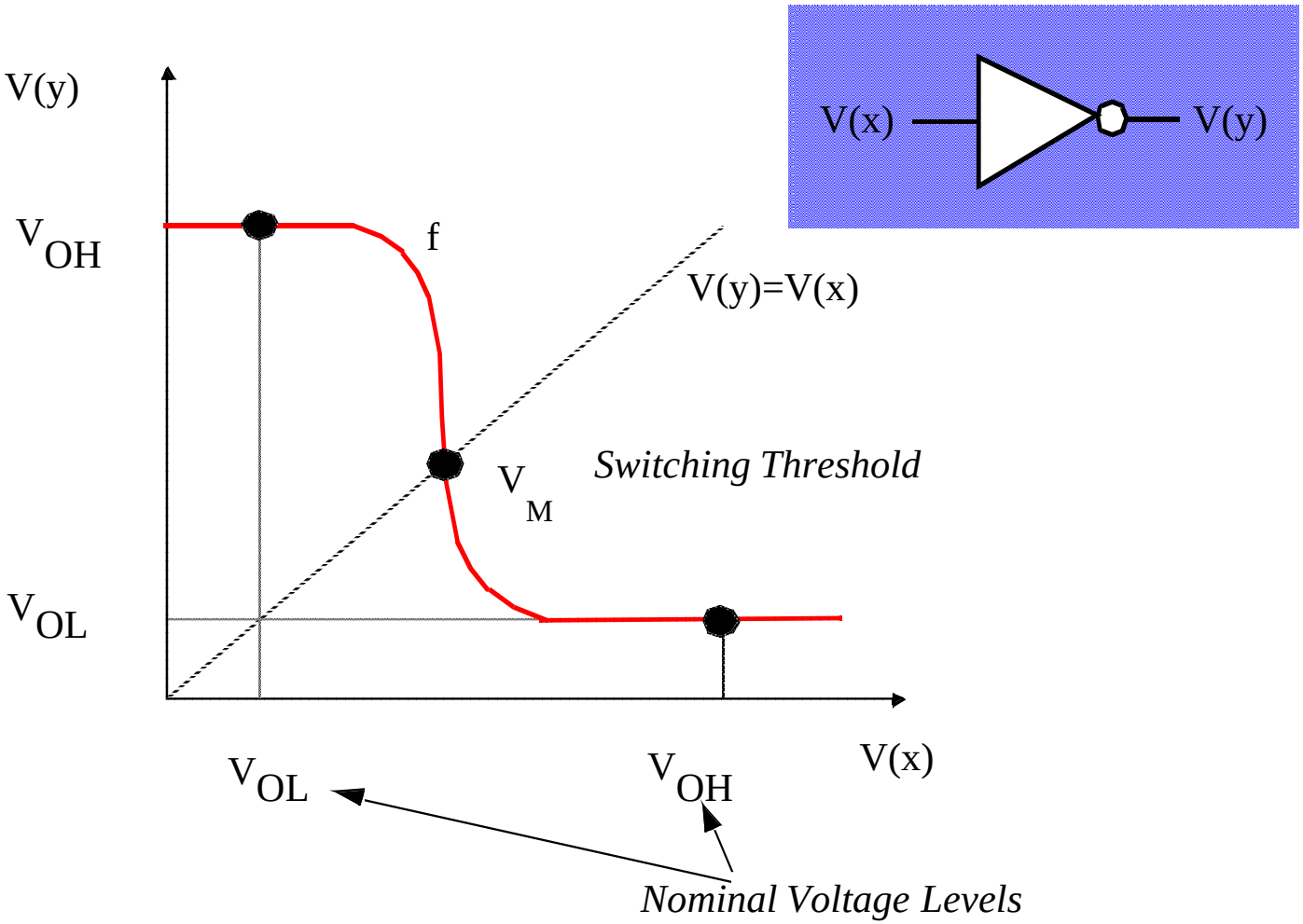
Low



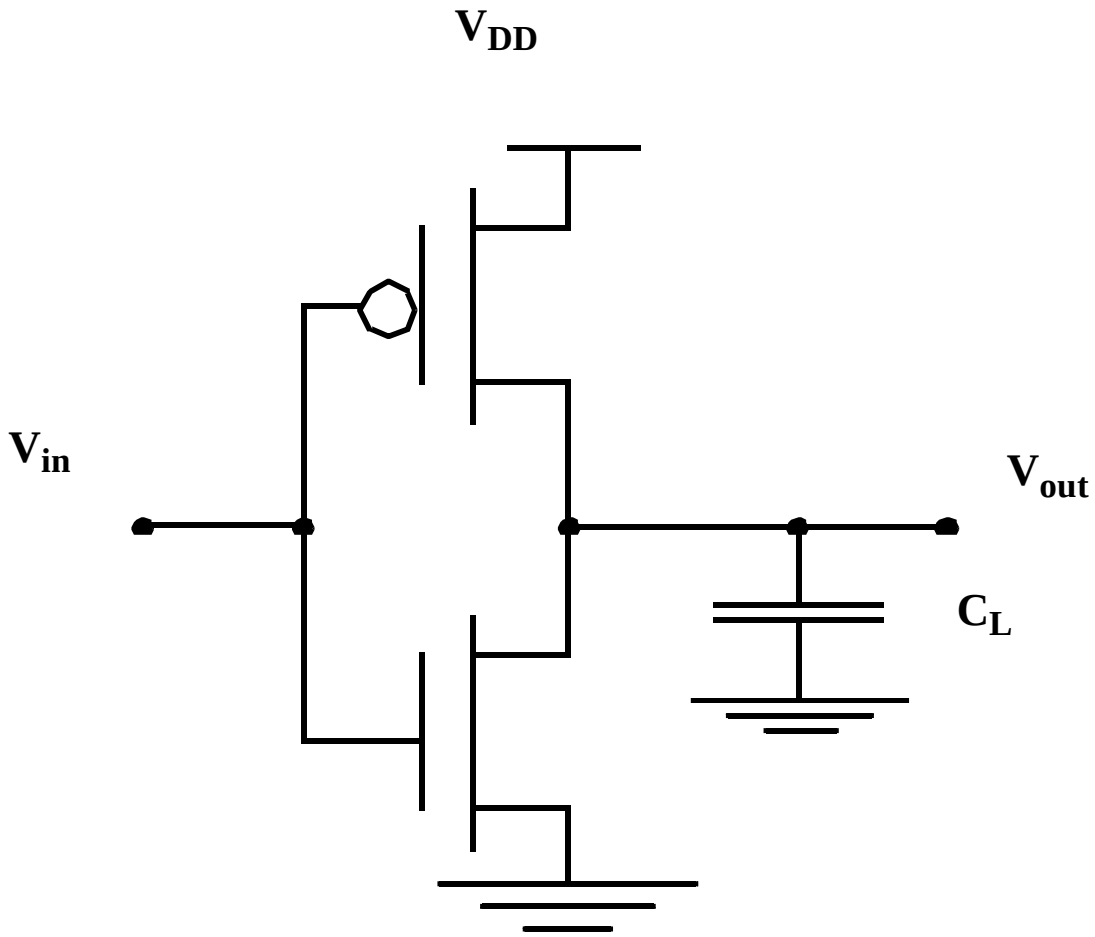
# DIGITAL GATES
# Fundamental Parameters

- Functionality
- Reliability, Robustness
- Area
- Performance
  - Speed (delay)
  - Power Consumption
  - Energy

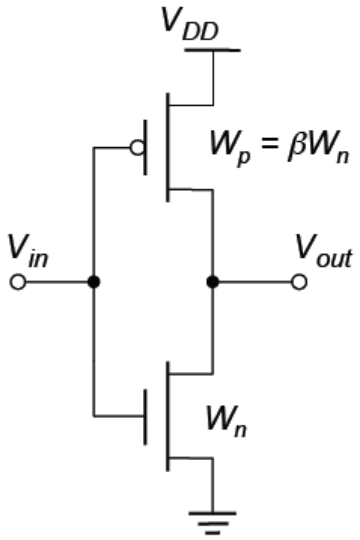
# DC Operation: Voltage Transfer Characteristic



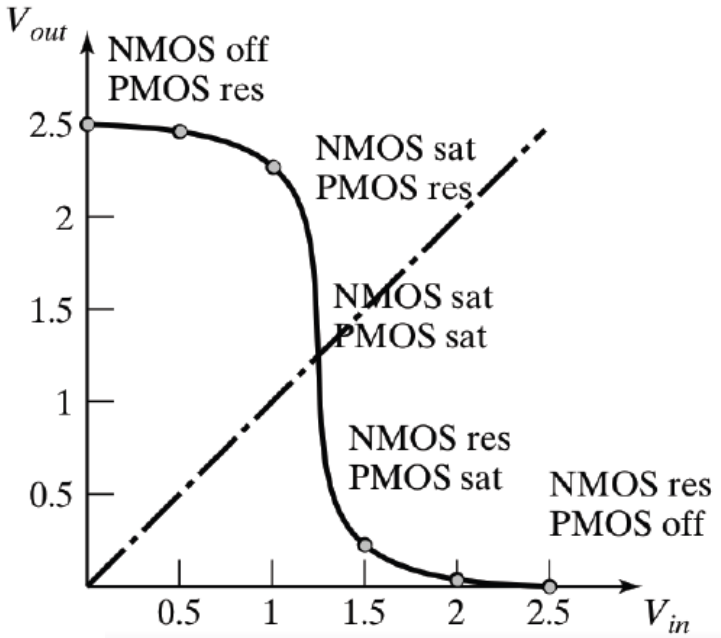
# The CMOS Inverter: A First Glance



# The CMOS Inverter




# CMOS Inverter VTC




- **Cut-off :** $I_{ds} = 0$ (for now)
  when $V_{gs} < V_T$

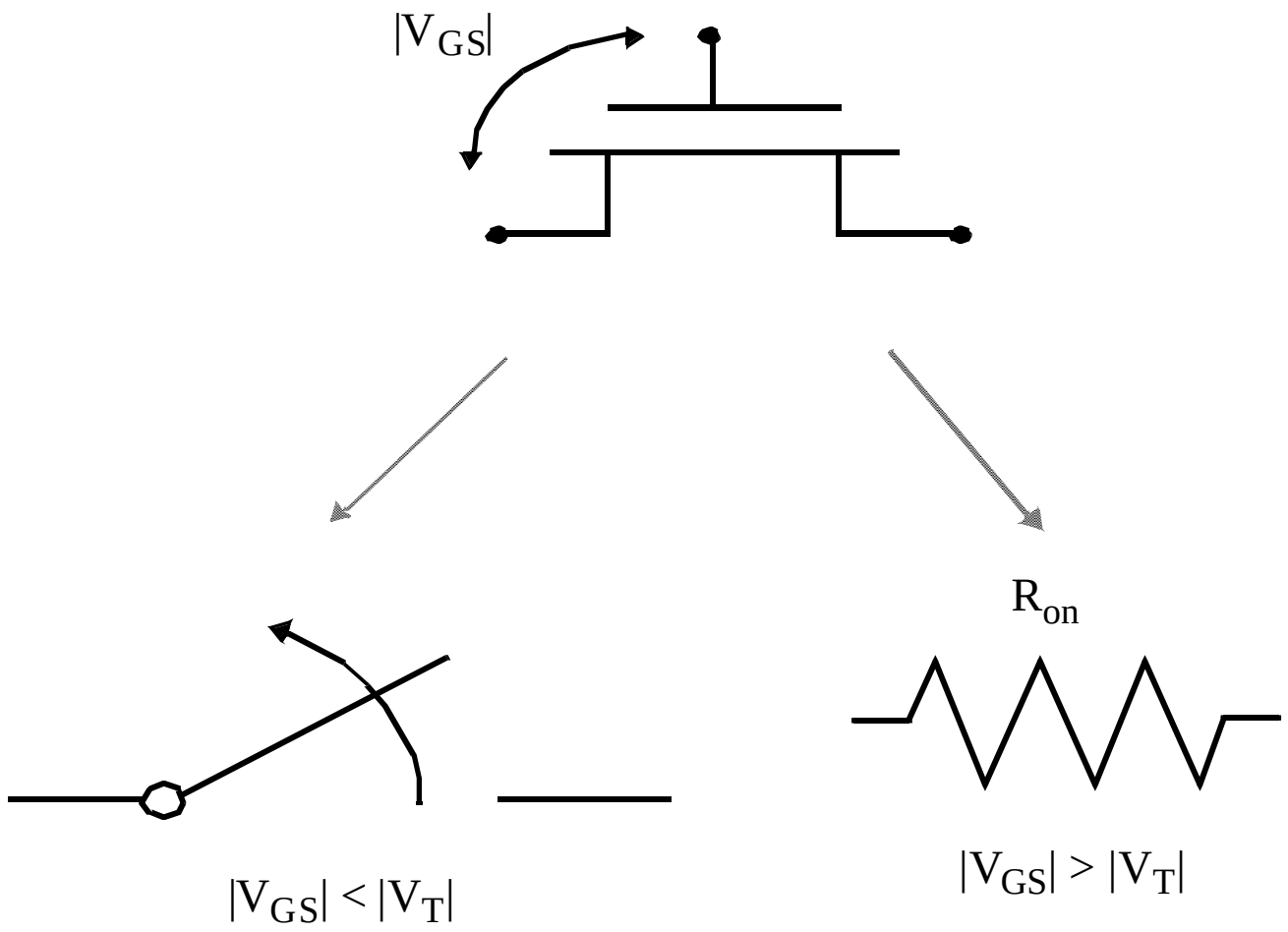
- **Linear :** $I_{ds} = \beta([V_{gs} - V_T]\ V_{ds} - \frac{V_{ds}^2}{2})$
  when $0 < V_{ds} < V_{gs} - V_T$

- **Saturation :** $I_{ds} = \frac{\beta}{2}(V_{gs} - V_T)^2$ (for now)
  when $0 < V_{gs} - V_T < V_{ds}$
  This is obtained by using $V_{ds} = V_{gs} - V_T$ in the equation for linear $I_{ds}$ (see comment two pages prior to this one)

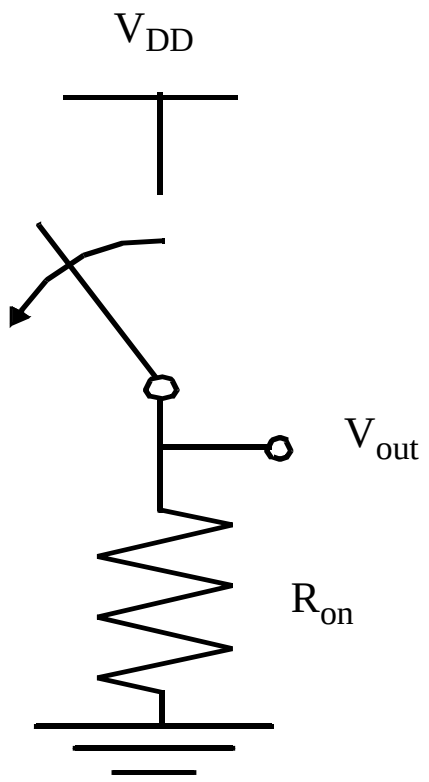
- Where $\beta = \frac{\mu\varepsilon}{t_{ox}}(\frac{W}{L})$

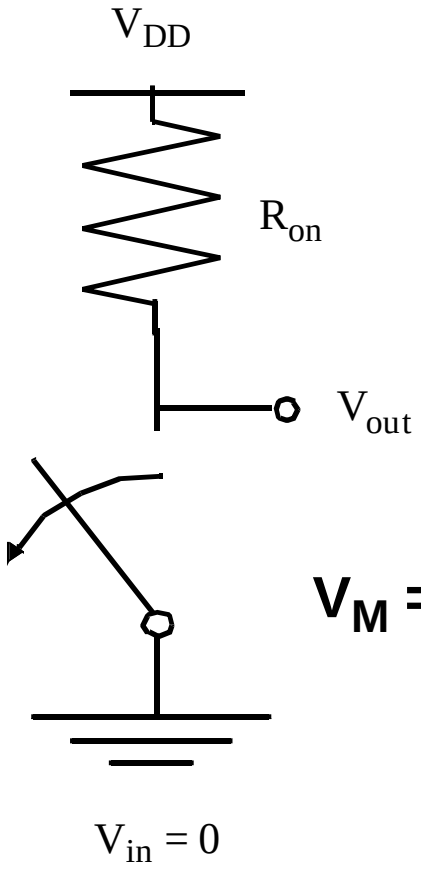
- **Cut-off :** $V_{gs} < V_T$
- **Linear :** $0 < V_{ds} < V_{gs} - V_T$
- **Saturation :** $0 < V_{gs} - V_T < V_{ds}$

# Switch Model of CMOS Transistor



# CMOS Inverter: Steady State Response



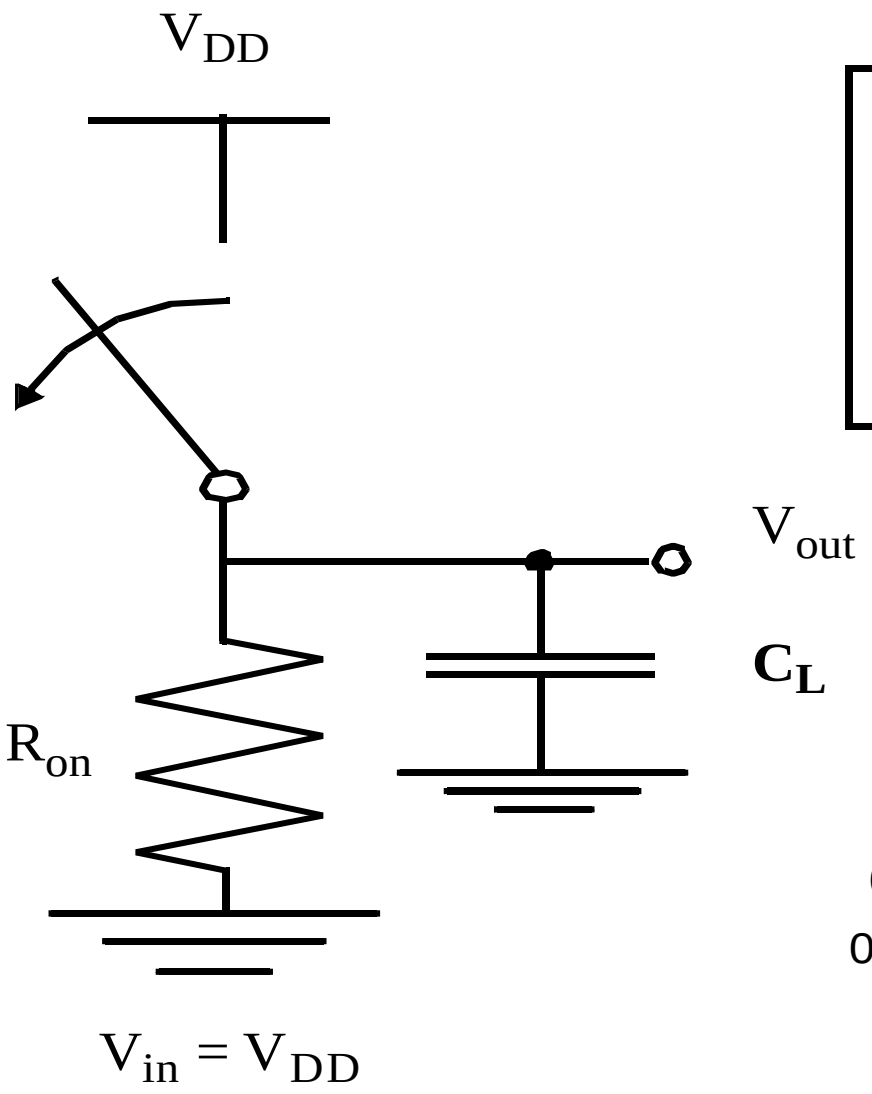




$V_{OH} = V_{DD}$

$V_{OL} = 0$

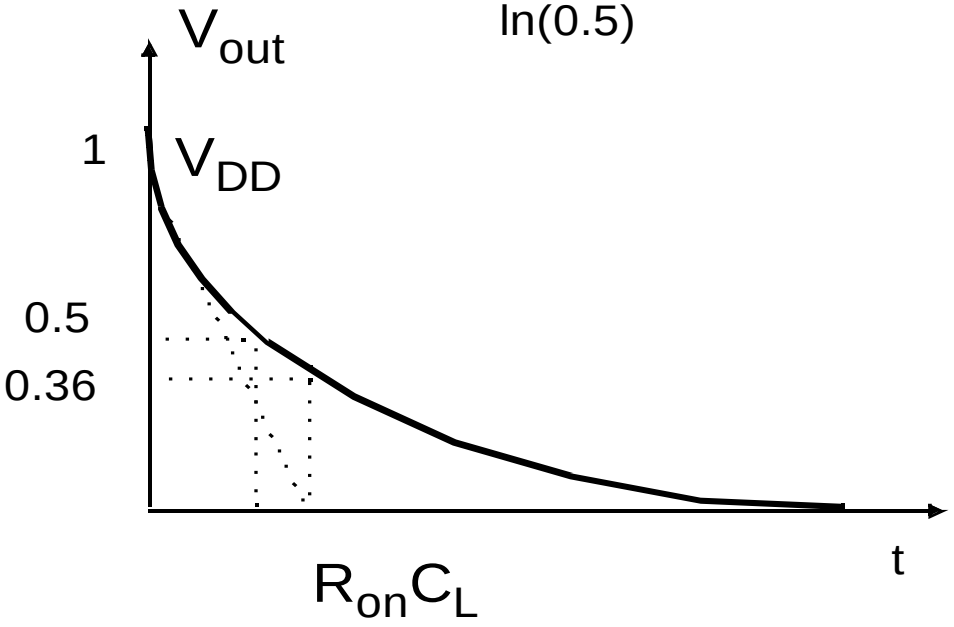
$V_M = f(R_{onn}, R_{onp})$

# CMOS Inverter: Transient Response

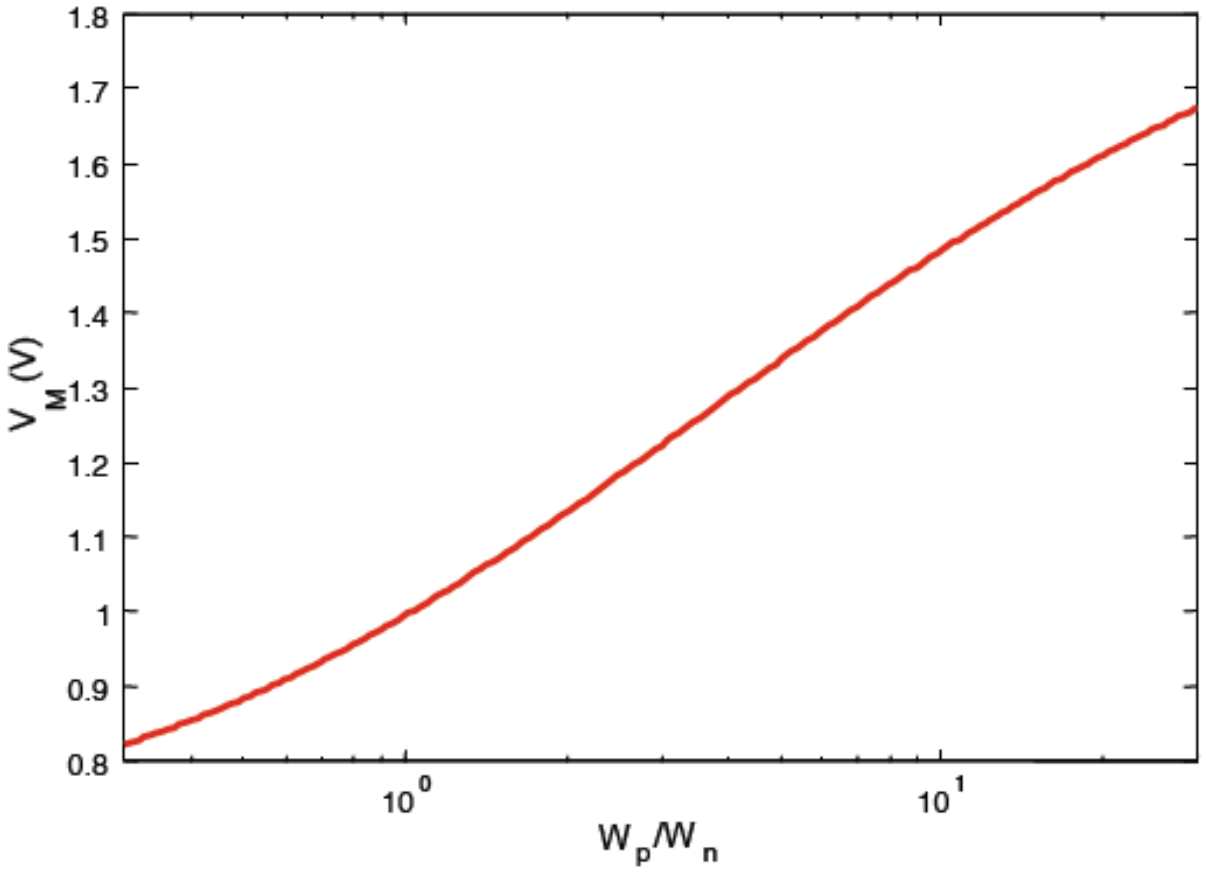



$$t_{pHL} = f(R_{on}.C_L)$$
$$= 0.69\ R_{on}C_L$$

ln(0.5)

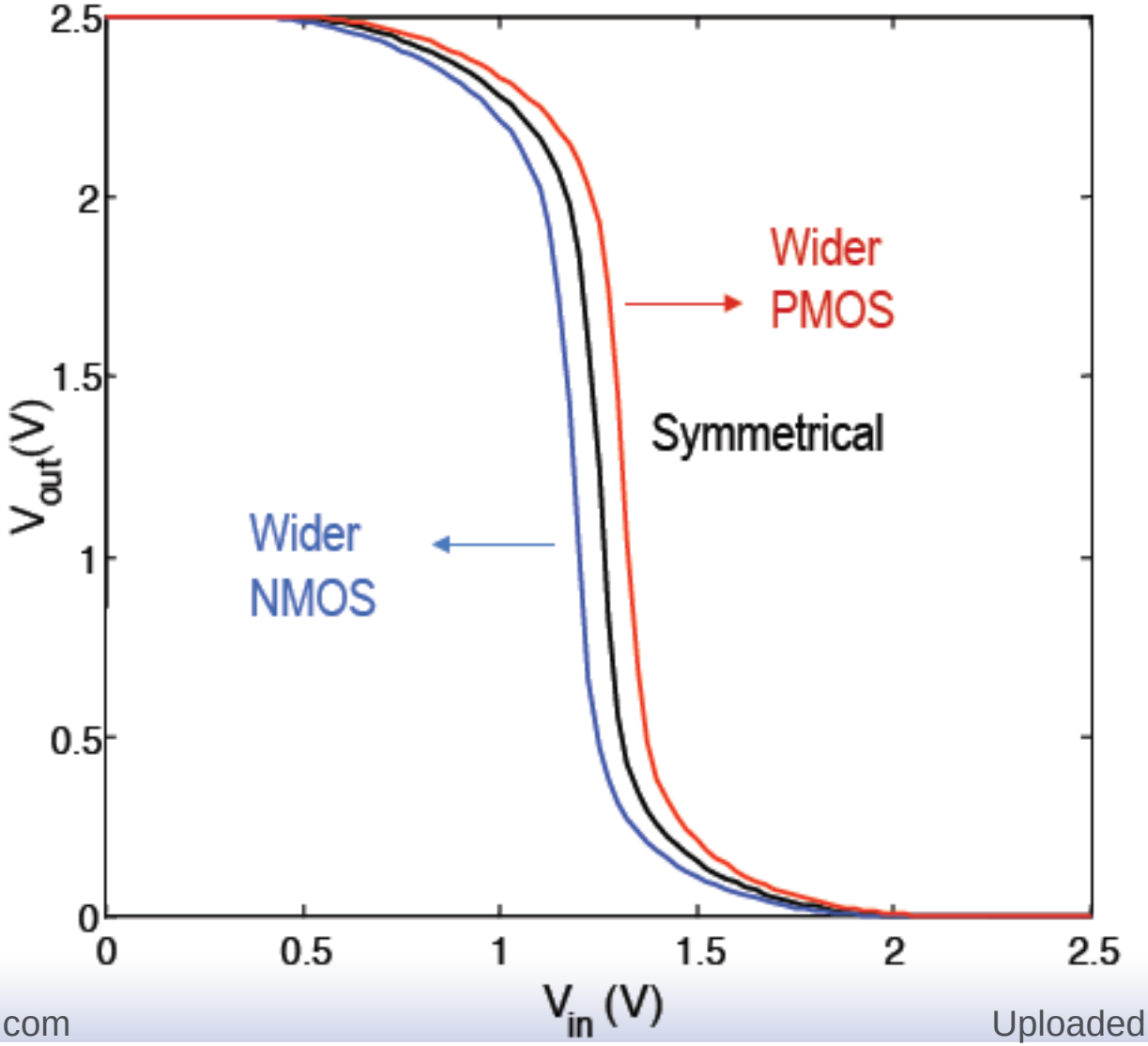


# Switching Threshold as a Function of Transistor Ratio

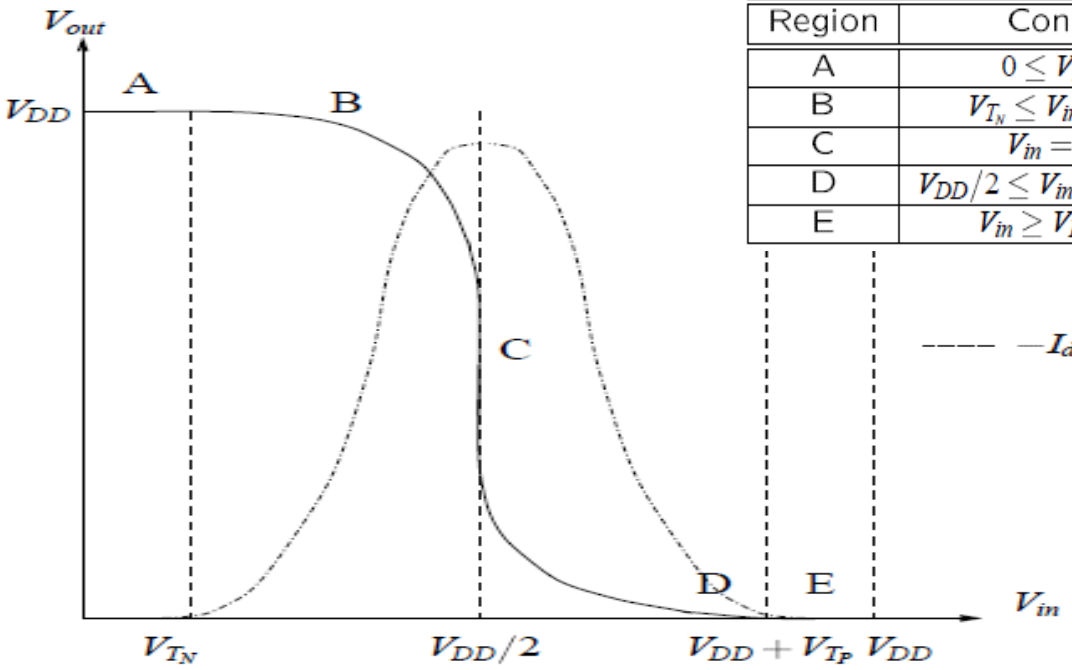



$$I_{dn}(V_M) = I_{dp}(V_M)$$

# Impact of Sizing



# DC inverter Characteristics




| Region | Condition | PMOS | NMOS | $V_{out}$ |
|---|---|---|---|---|
| A | $0 \le V_{in} \le V_{T_N}$ | linear | cutoff | $V_{DD}$ |
| B | $V_{T_N} \le V_{in} \le V_{DD}/2$ | linear | sat | see below |
| C | $V_{in} = V_{DD}/2$ | sat | sat | $V_{out} \neq f(V_{in})$ |
| D | $V_{DD}/2 \le V_{in} \le V_{DD} - \lvert V_{T_P}\rvert$ | sat | linear | see below |
| E | $V_{in} \ge V_{DD} - \lvert V_{T_P}\rvert$ | cutoff | linear | 0 |

Regions A, B, C, D and E based on state of P and N devices

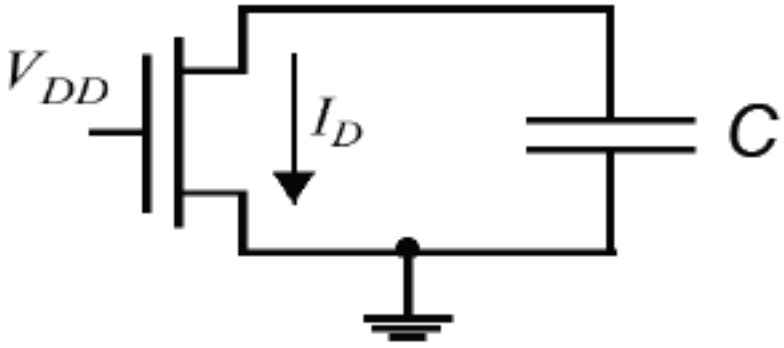
DC characteristics, so no capacitors involved. We will talk about capacitors when we consider AC characteristics.

- NMOS is easy to see, but how do we determine PMOS device state? For this, assume an inverter with $V_{DD} = 5V$, $V_{T_N} = -V_{T_P} = 1V$.

| $V_{in}$ | $V_{gs_P} - V_{T_P}$ | Relation | $V_{ds_P}$ | Device state |
|---|---|---|---|---|
| 0 | -5 + 1 | $<$ | 0 | linear |
| 1 | -4 + 1 | $<$ | $\sim -1$ | linear |
| 2.5 | -2.5 + 1 | $>$ | $\sim -2.5$ | saturation |
| 4 | -1 + 1 | $>$ | $\sim -4$ | saturation |
| 5 | 0 + 1 $\ge$ 0 | - | - | cutoff |

# MOS Transistor as a Switch

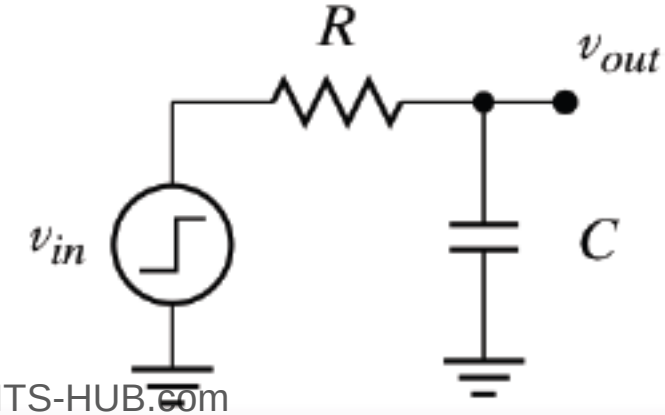
- Discharging a capacitor




$$i_D = i_D(v_{DS})$$

$$i_D = C\frac{dV_{DS}}{dt}$$

- We modeled this with:



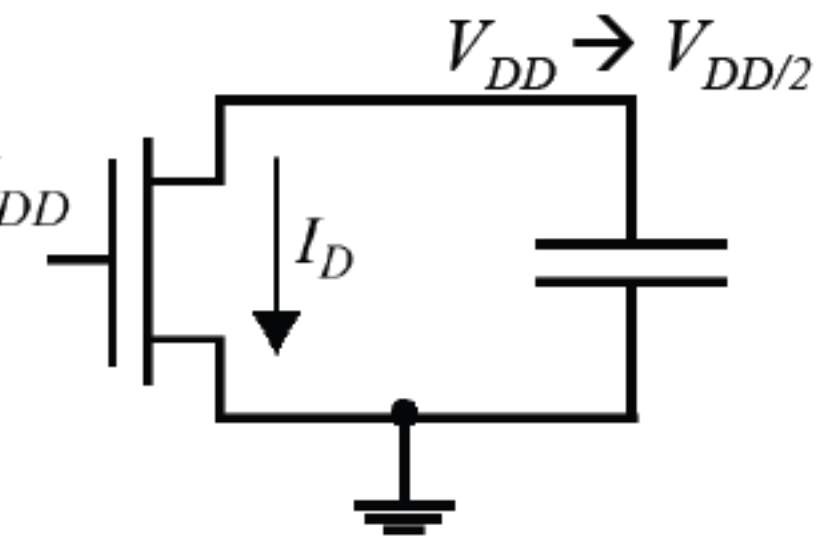

$$t_p = \ln(2)\,RC$$

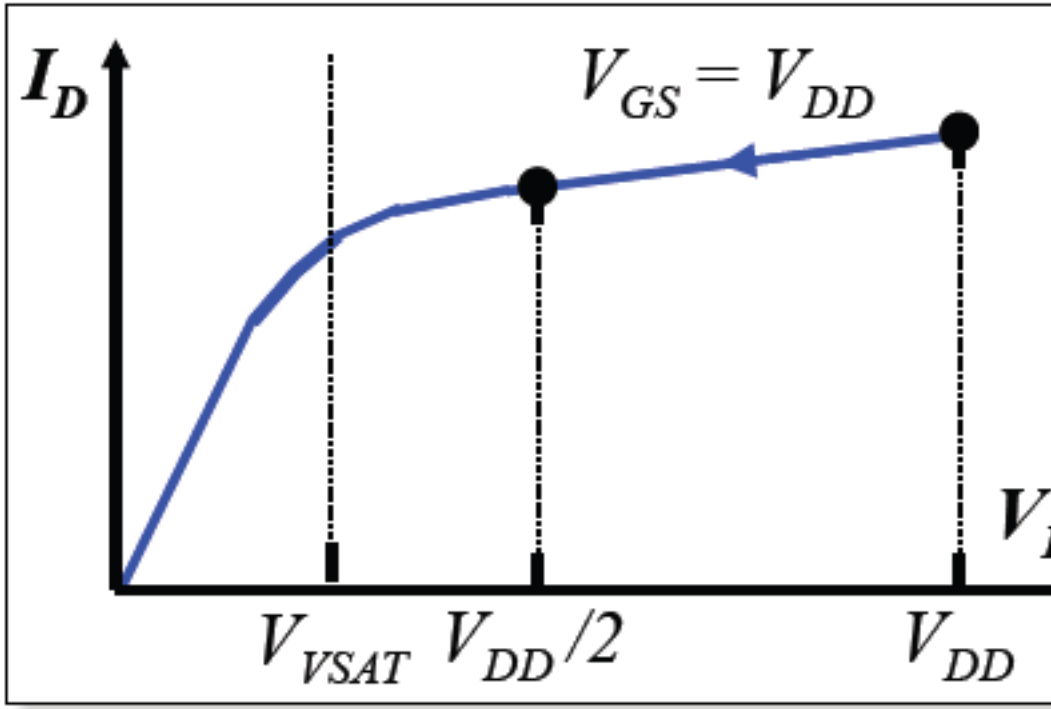
# MOS Transistor as a Switch

- Real transistors aren't exactly resistors
  - Look more like current sources in saturation
- Two questions:
  - Which region of IV curve determines delay?
  - How can that match up with the RC model?

# Transistor Discharging a Capacitor

- With a step input:

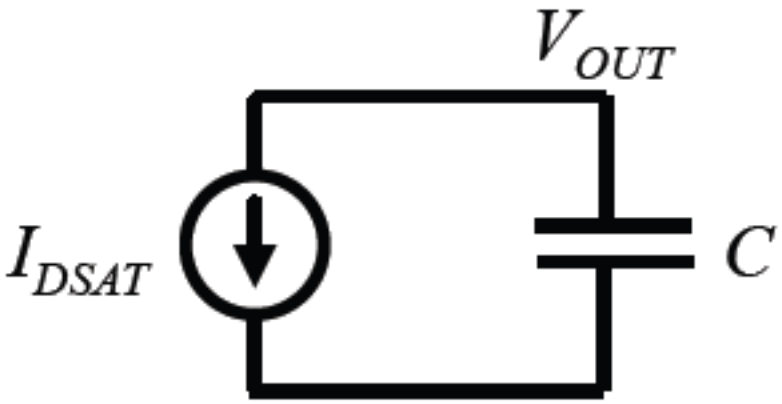


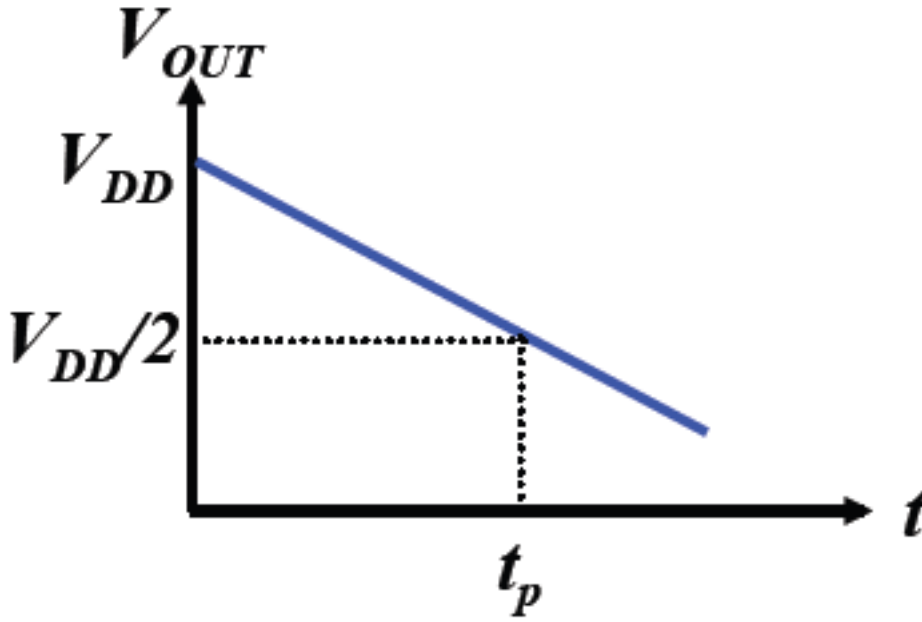




Transistor is in (velocity) saturation during entire transition from $V_{DD}$ to $V_{DD}/2$

# Switching Delay

- In saturation, transistor basically acts like a current sourc

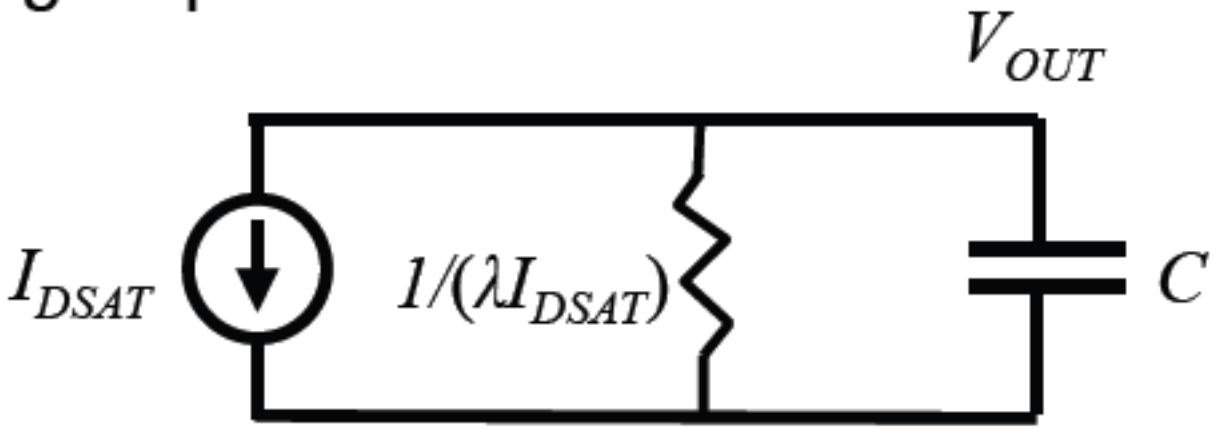



$$V_{OUT} = V_{DD} - (I_{DSAT}/C)t \longrightarrow \boxed{t_p = C(V_{DD}/2)/I_{DSAT}}$$

# Switching Delay (with Output Conductan

- Including output conductance:

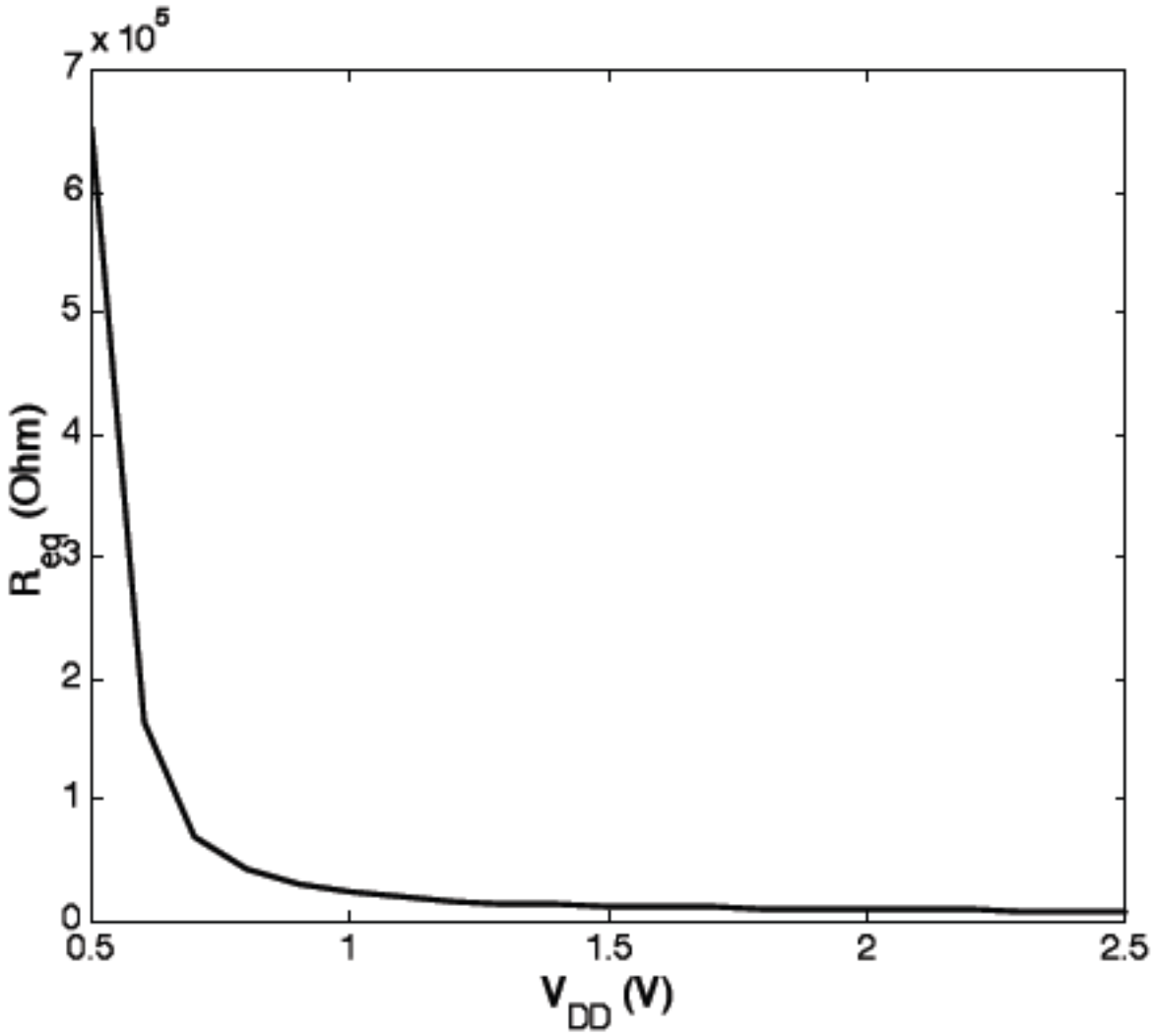



$$V_{OUT} = \left(V_{DD} + \lambda^{-1}\right) e^{-t/(C/\lambda I_{DSAT})} - \lambda^{-1}$$

- For "small" λ:

$$t_p \approx \frac{C\left(V_{DD}/2\right)}{\left(1+\lambda V_{DD}\right) I_{DSAT}}$$

# The Transistor as a Switch

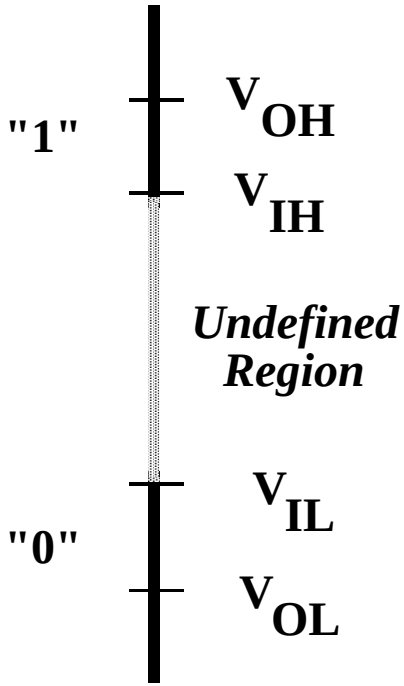


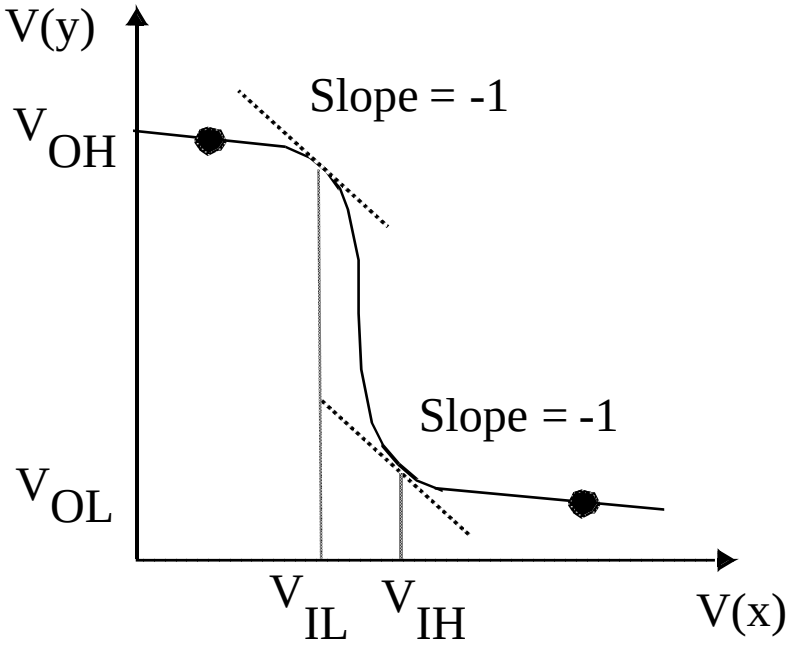
# The Transistor as a Switch

**Table 3.3** Equivalent resistance $R_{eq}$ ($W/L$= 1) of NMOS and PMOS transistors in 0.25 μm CMOS process (with $L = L_{min}$). For larger devices, divide $R_{eq}$ by $W/L$.

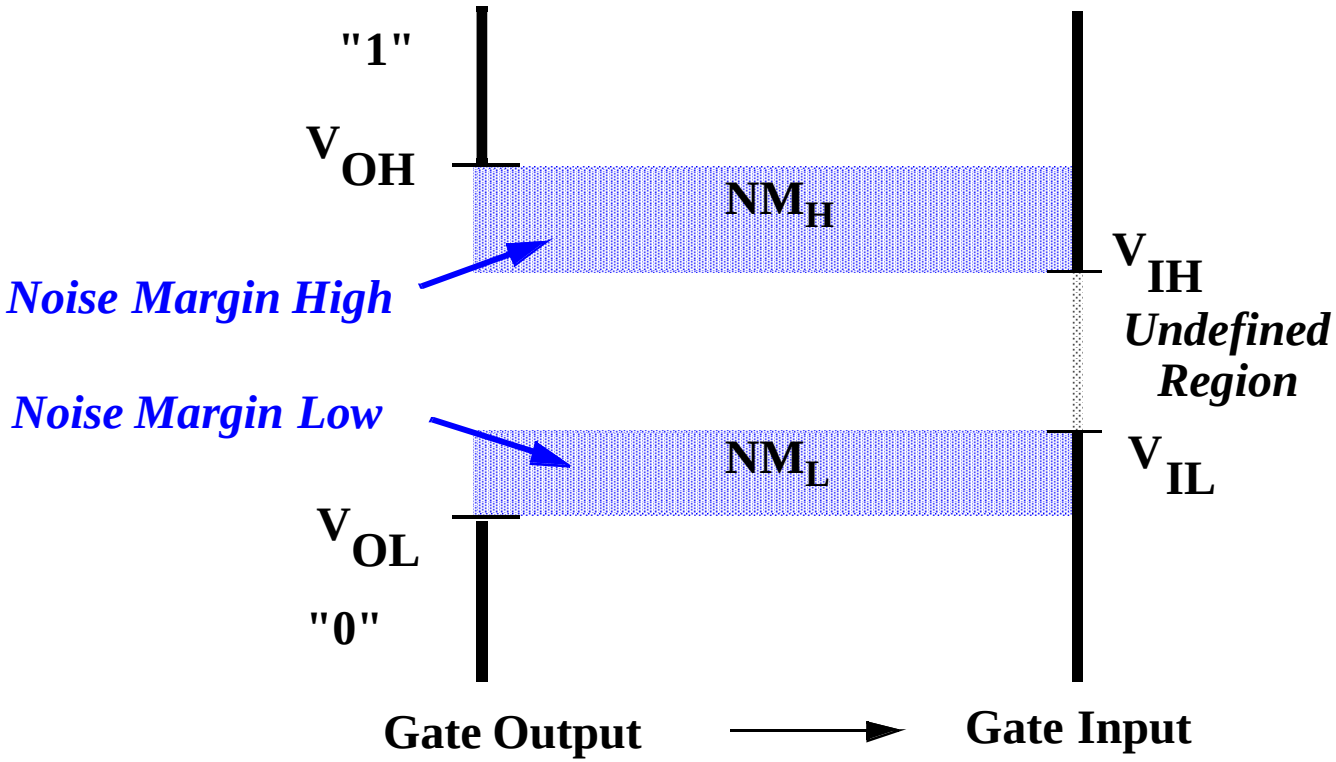
| $V_{DD}$ (V) | 1 | 1.5 | 2 | 2.5 |
|---|---|---|---|---|
| NMOS (kΩ) | 35 | 19 | 15 | 13 |
| PMOS (kΩ) | 115 | 55 | 38 | 31 |

# Mapping between analog and digital signals



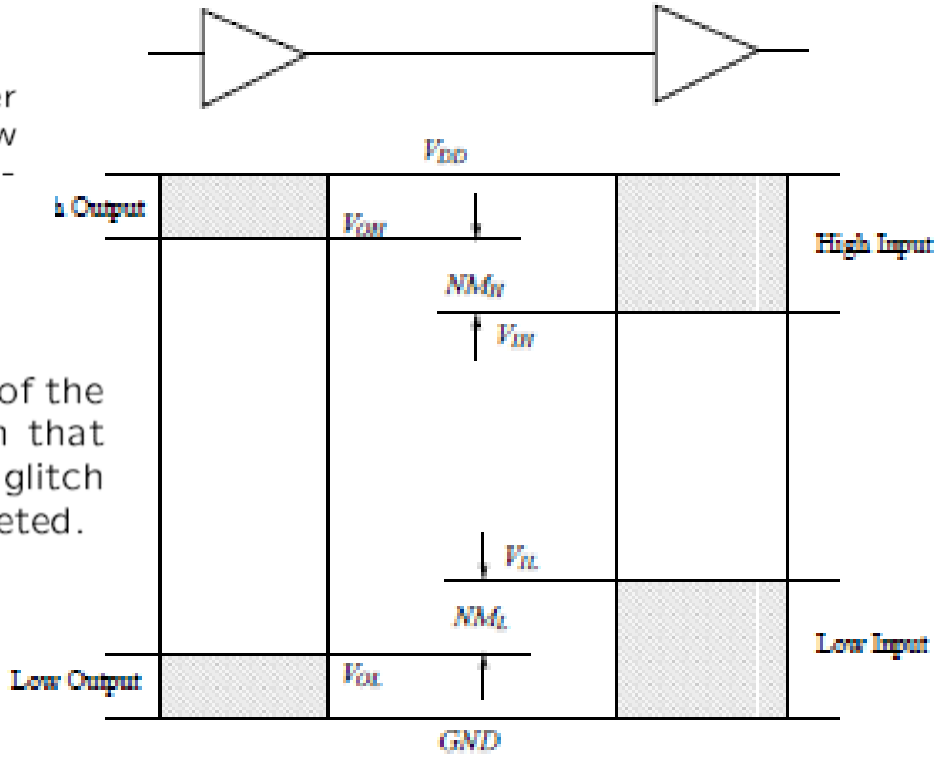



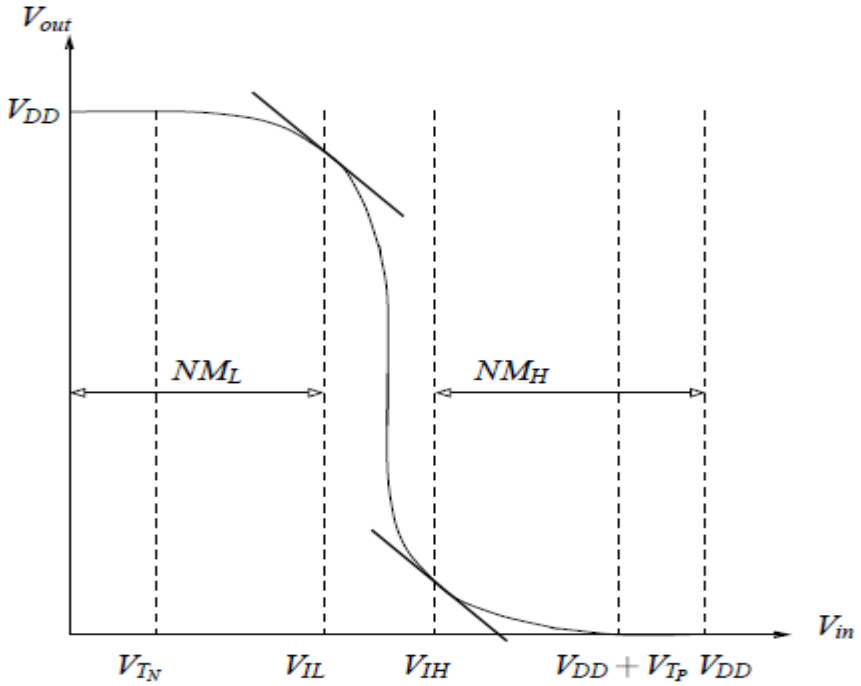
# Definition of Noise Margins

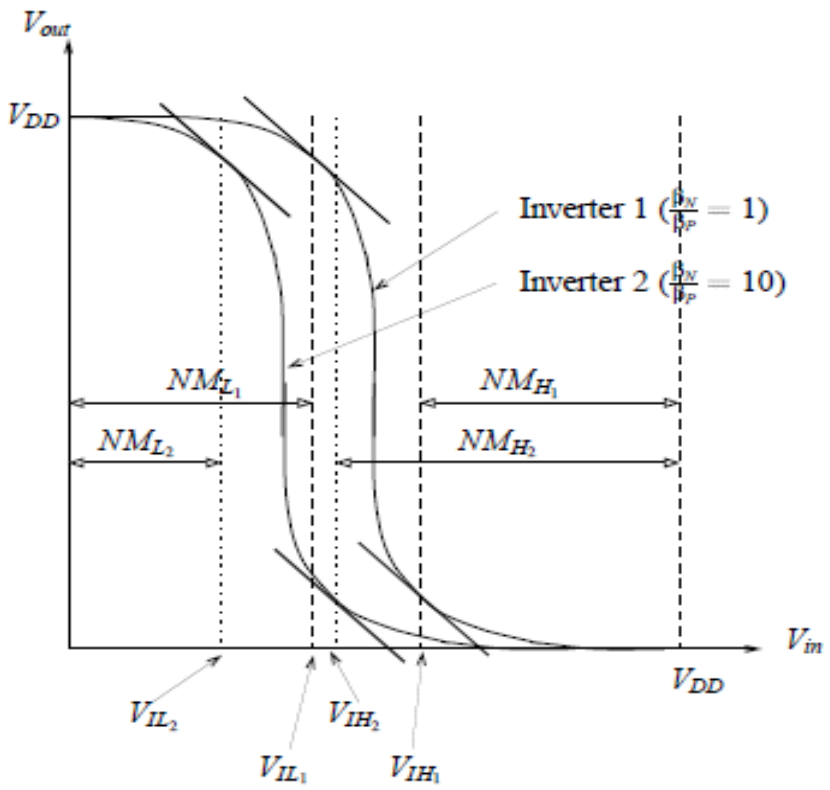


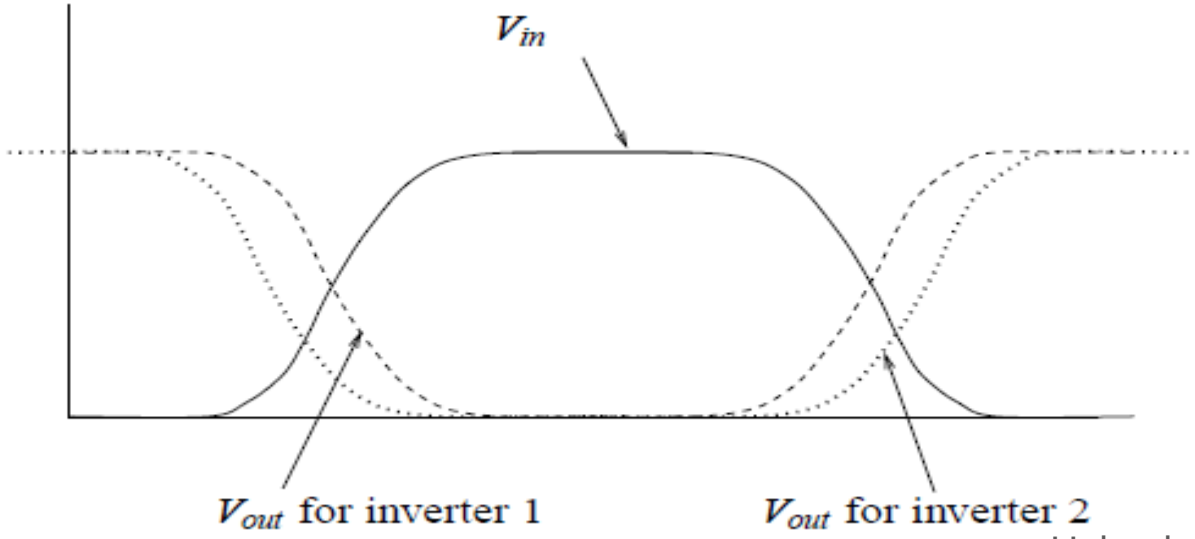
# Noise Margin

- $NM_H$ and $NM_L$ are the high-side and low-side noise margins.
- The high output excursion should **not** be larger than the high input excursion. Same for the low excursions. If this is violated, then the corresponding noise margin is negative.
- Why worry about it? Perhaps the $V_{DD}$ or $GND$ of the driver glitches relative to the driven gate. In that situation, I want to know what magnitude of glitch can I tolerate before a wrong value is interpreted.

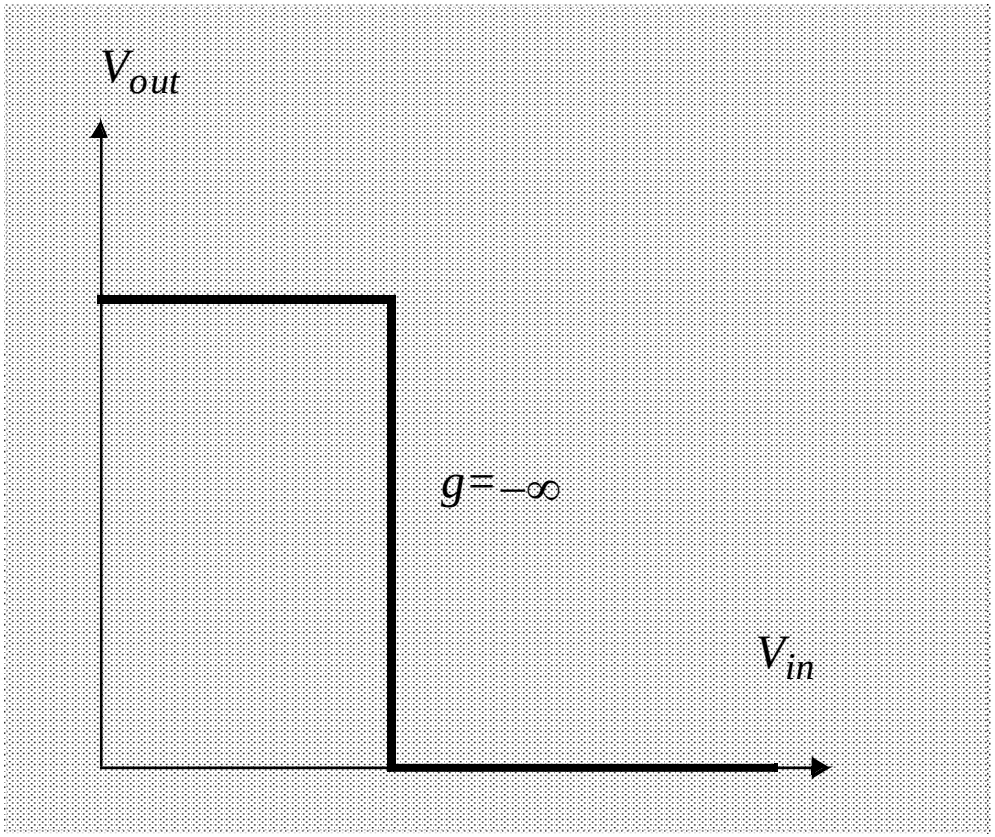




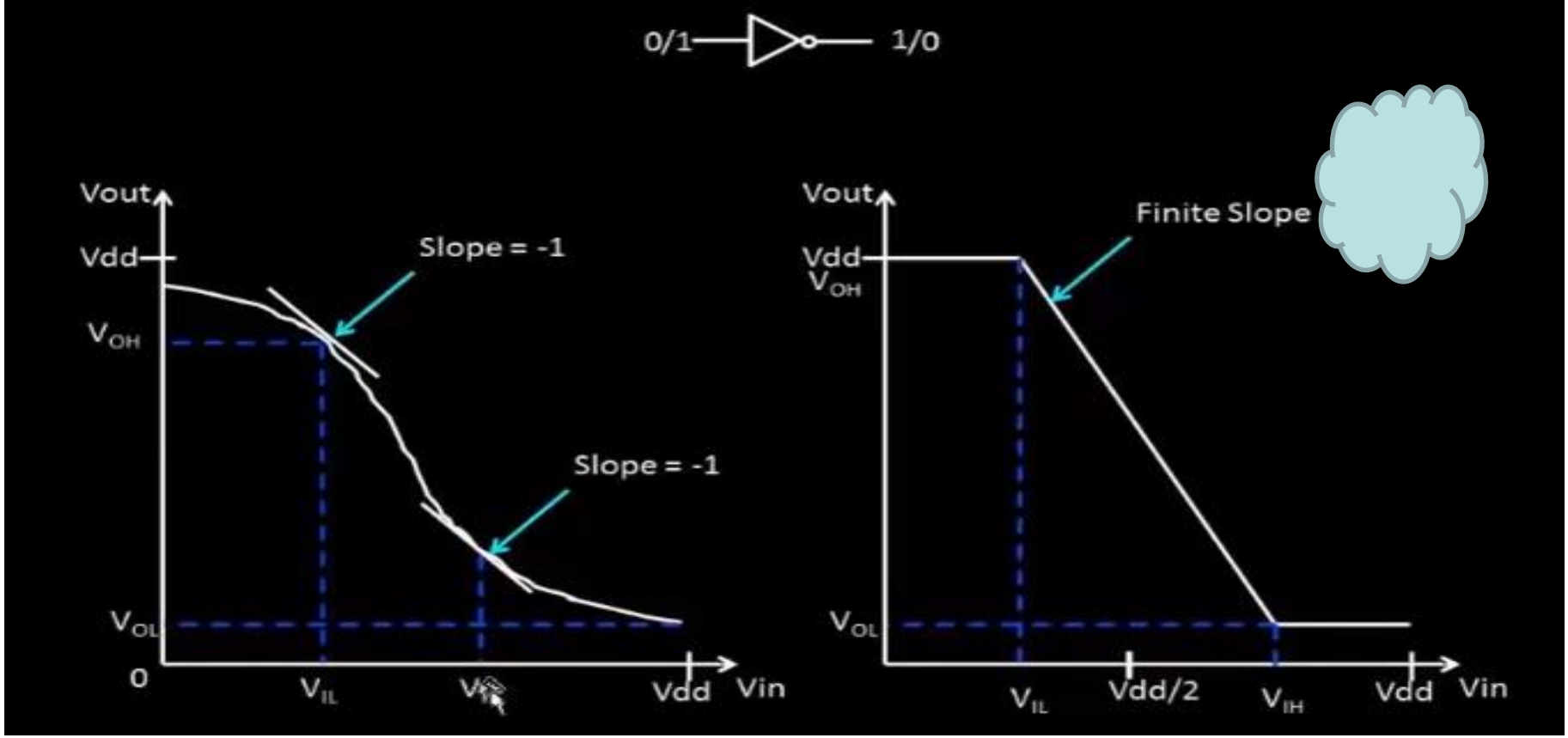



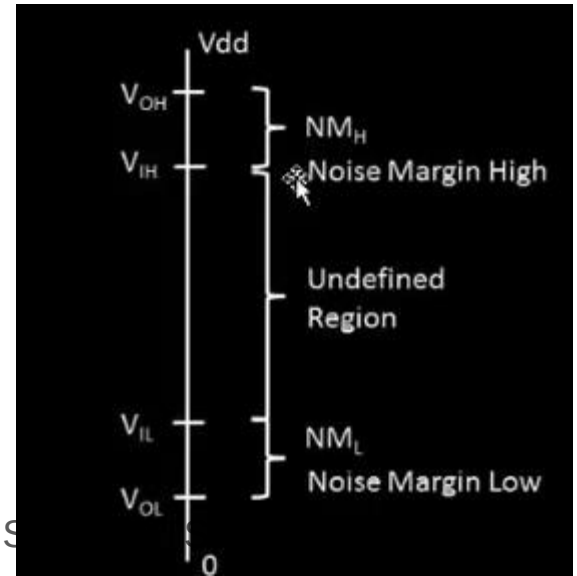
# The Ideal Gate

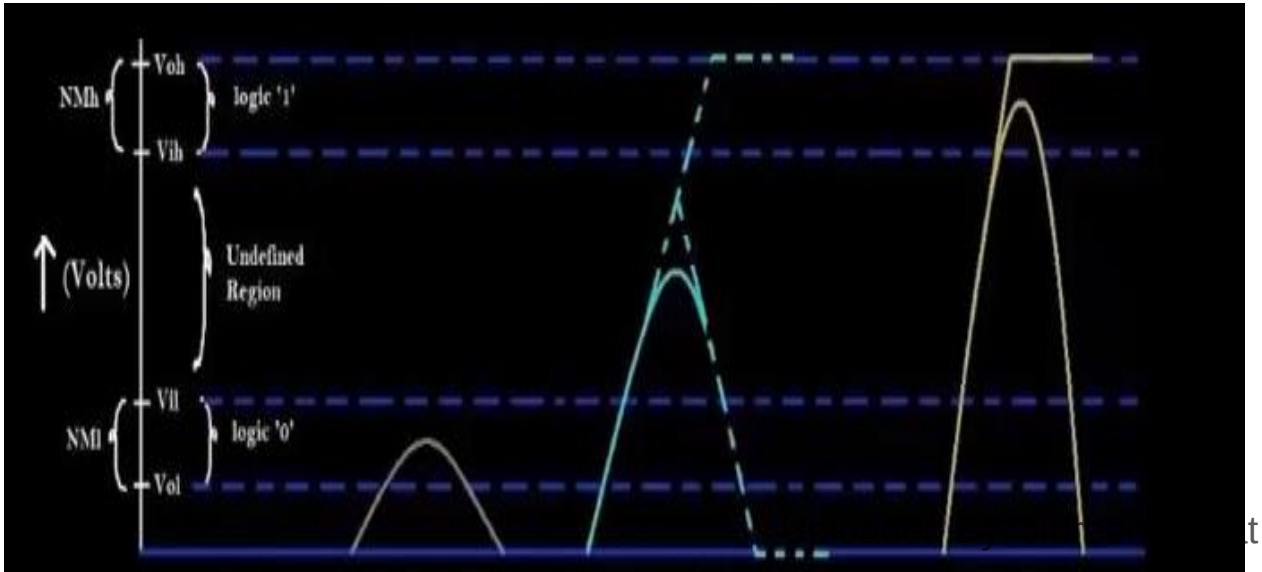



$R_i = \infty$

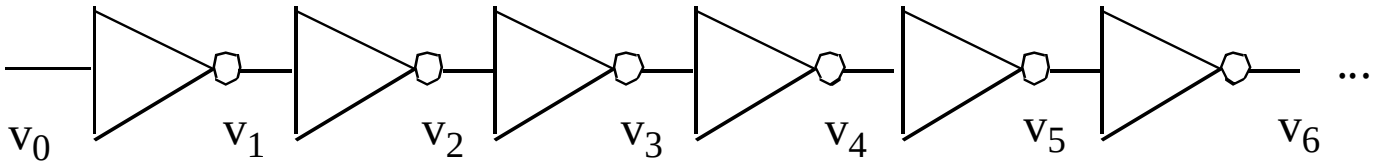
$R_o = 0$


0/1
1/0
Vout
Vdd
$V_{OH}$
Slope = -1
Slope = -1
$V_{OL}$
0
$V_{IL}$
Vdd
Vin
Vout
Vdd
$V_{OH}$
Finite Slope
$V_{OL}$
$V_{IL}$
Vdd/2
$V_{IH}$
Vdd
Vin

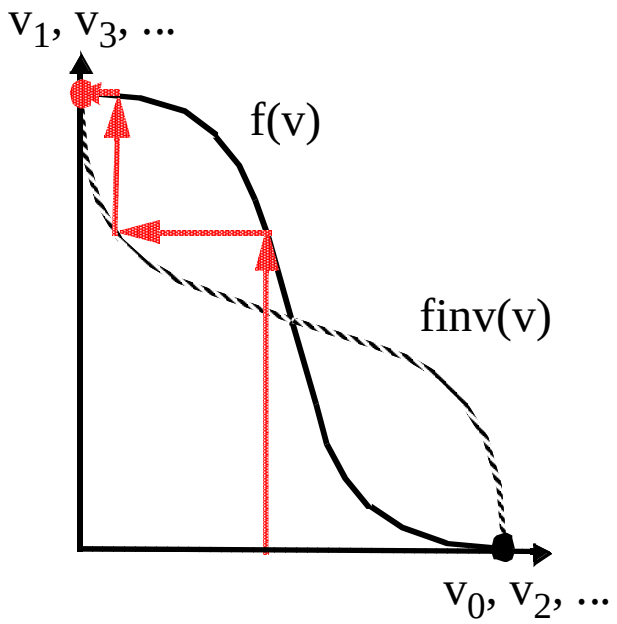

Vdd
$V_{OH}$
$NM_H$
$V_{IH}$
Noise Margin High
Undefined
Region
$V_{IL}$
$NM_L$
Noise Margin Low
$V_{OL}$
0


NMh
Voh
logic '1'
Vih
(Volts)
Undefined
Region
Vil
NMl
logic '0'
Vol

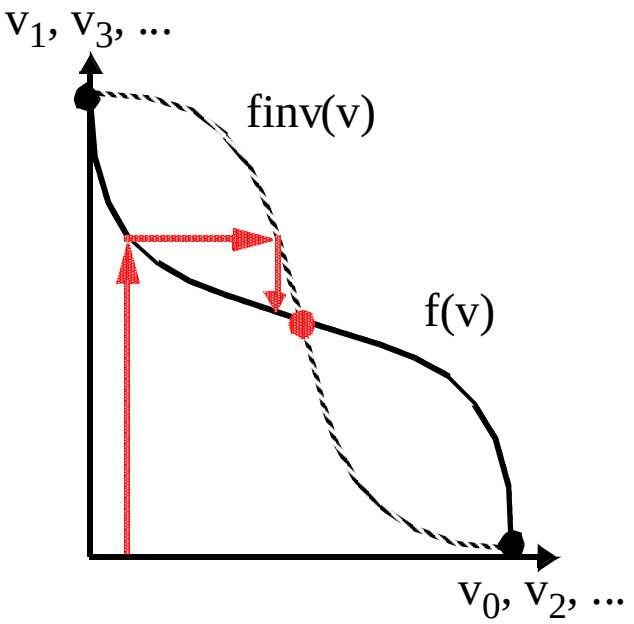
# The Regenerative Property




(a) A chain of inverters.

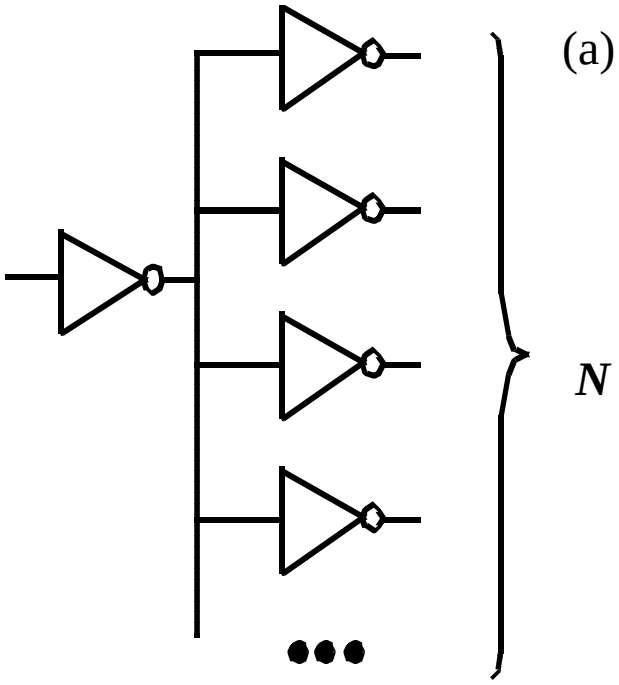



(b) Regenerative gate

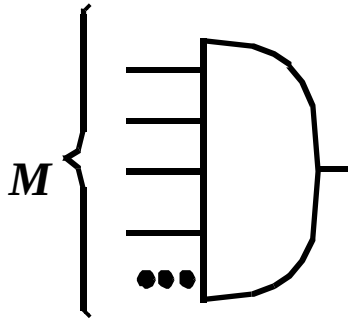



(c) Non-regenerative gate

# Fan-in and Fan-out

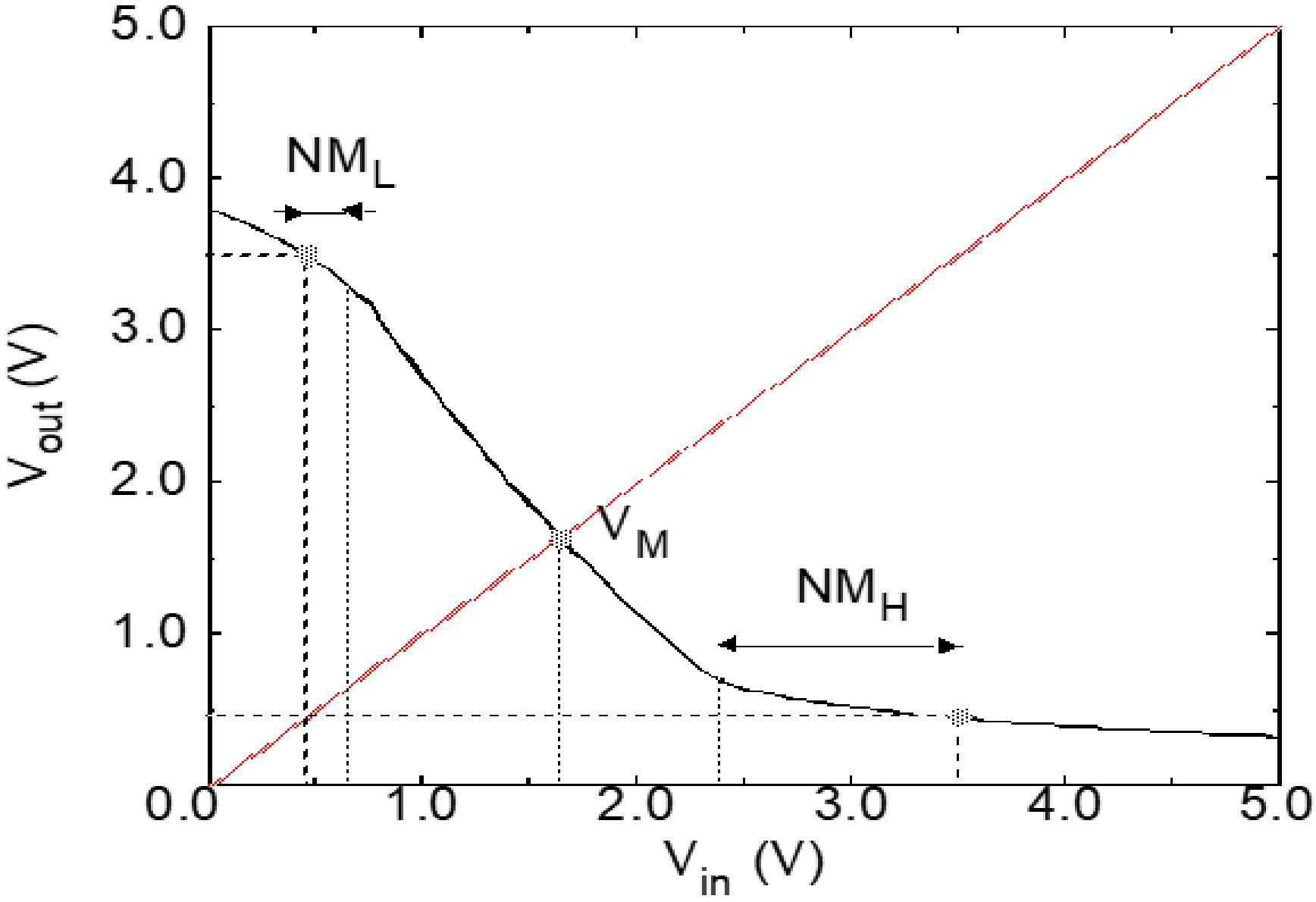



(a) Fan-out *N*

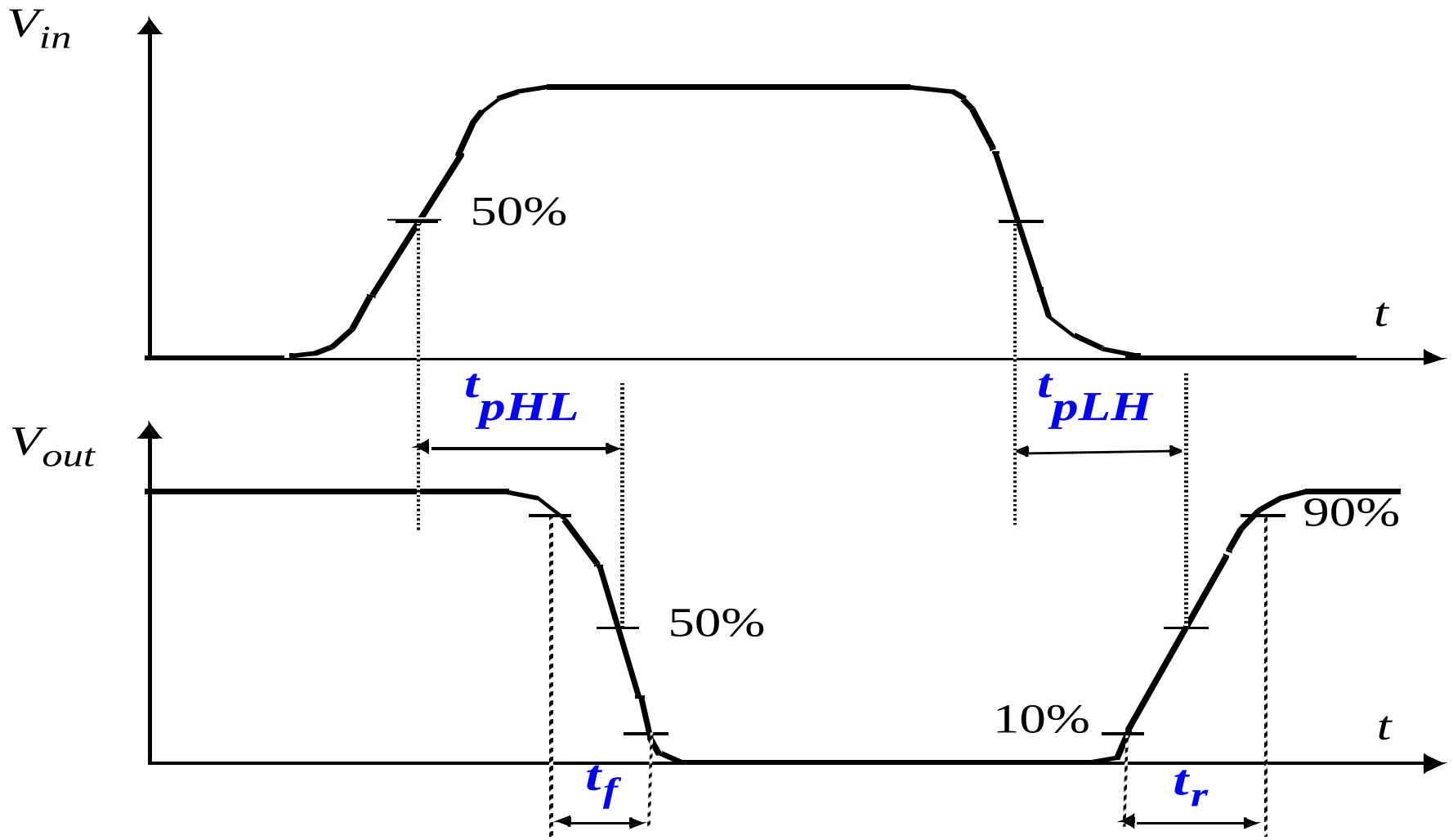



(b) Fan-in *M*

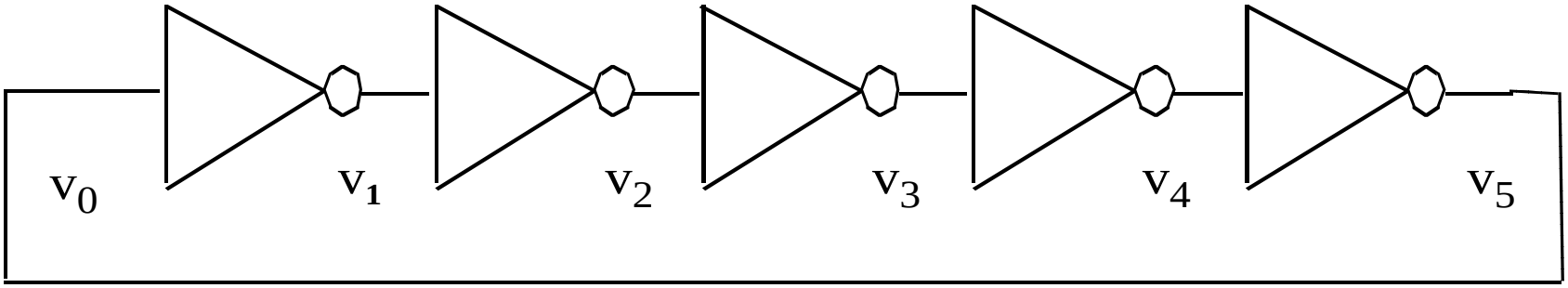
# VTC of Real Inverter

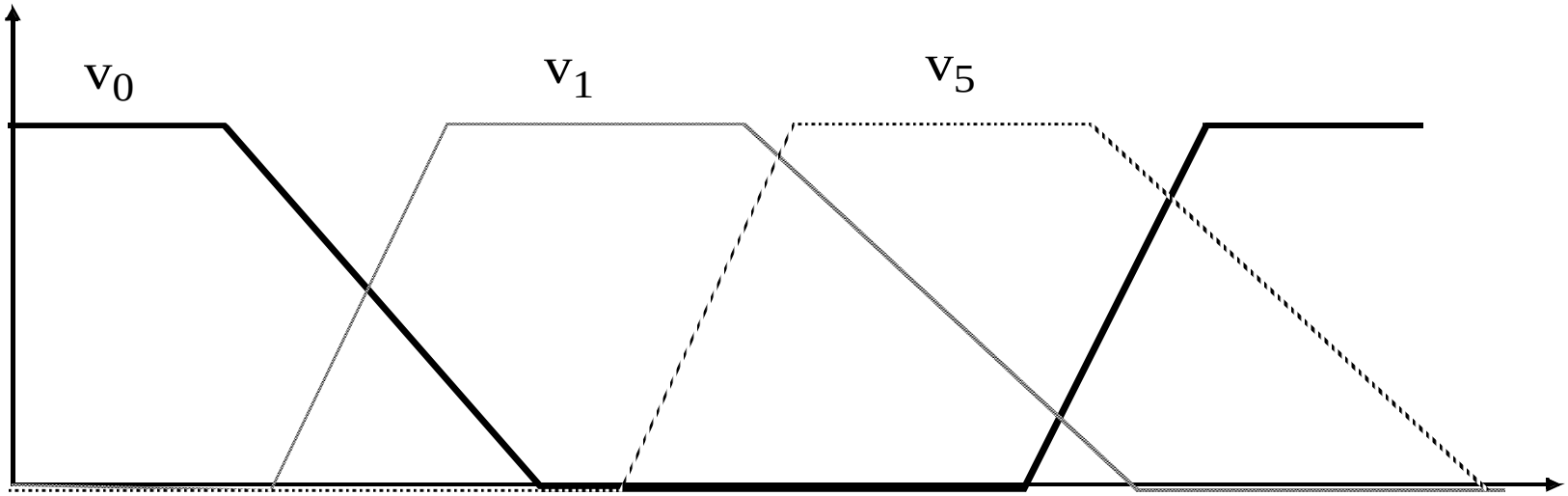


# Delay Definitions



# Ring Oscillator




$$T = 2 \times t_p \times N$$

# Power Dissipation

$$P_{peak} = i_{peak} V_{supply} = max(p(t))$$

$$P_{av} = \frac{1}{T}\int_0^T p(t)dt = \frac{V_{supply}}{T}\int_0^T i_{supply}(t)dt$$

**Power-Delay Product**

$$PDP = t_p \times P_{av}$$

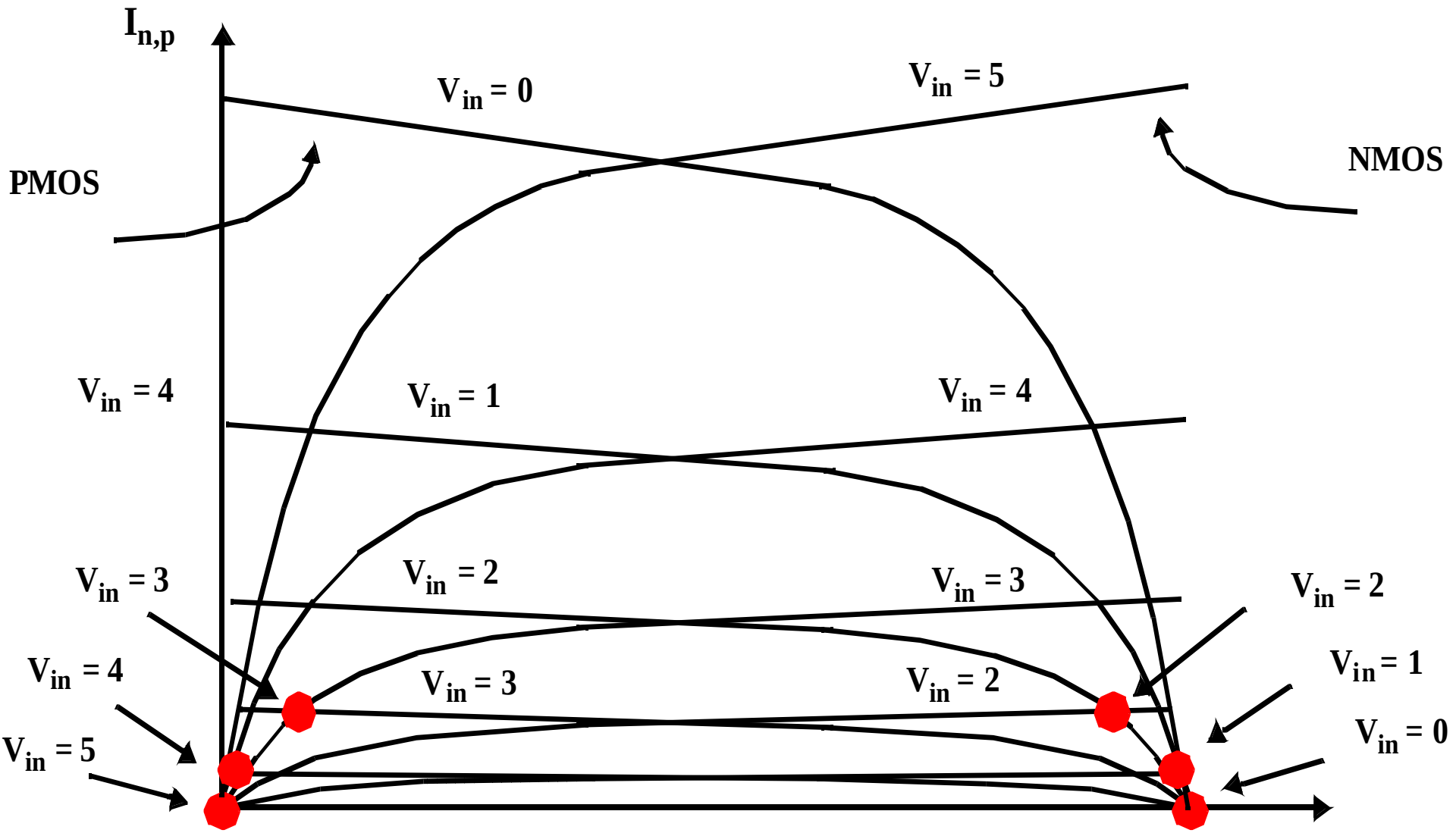
= Energy dissipated per operation

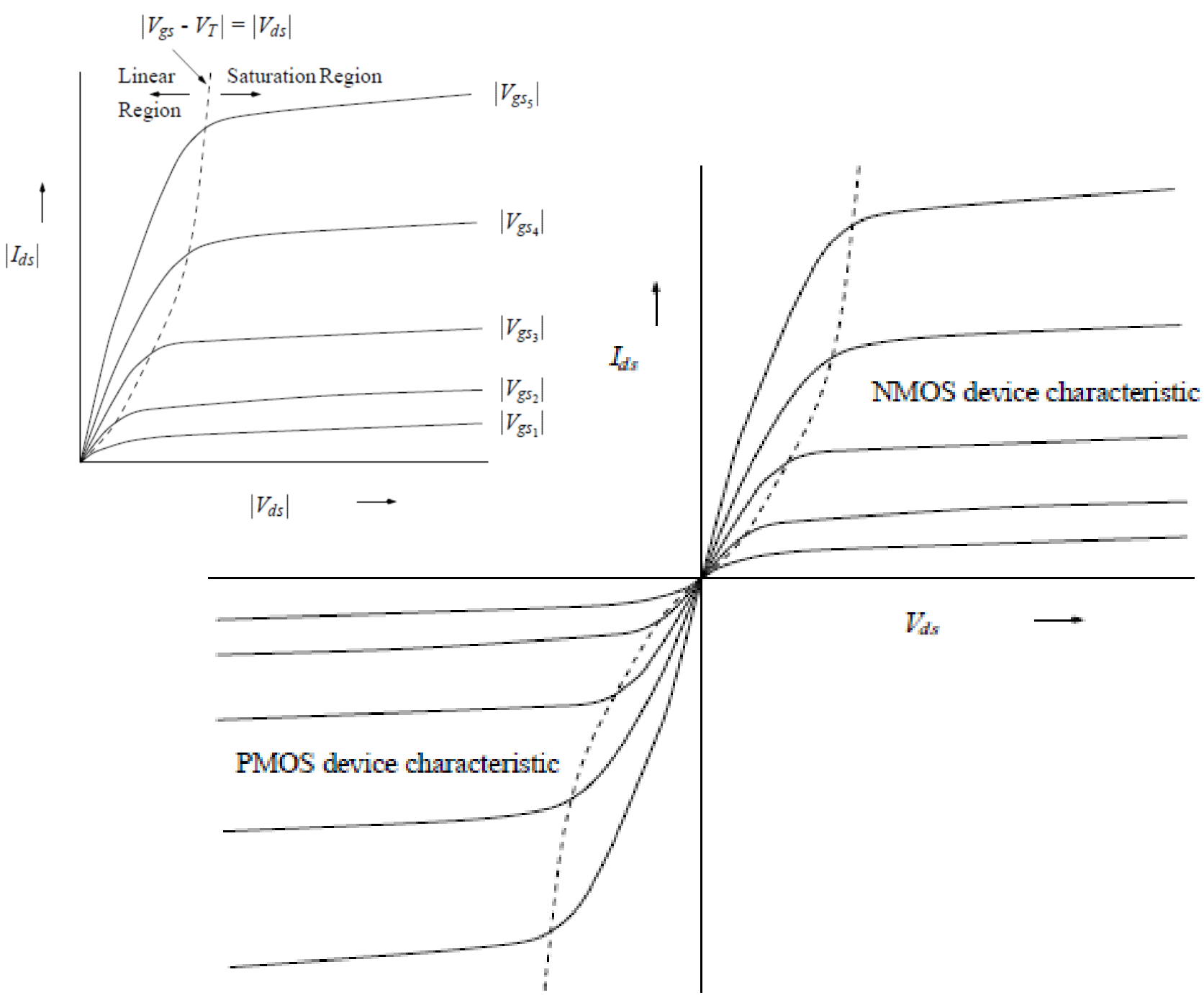
# CMOS Properties

- Full rail-to-rail swing
- Symmetrical VTC
- Propagation delay function of load capacitance and resistance of transistors
- No static power dissipation
- Direct path current during switching

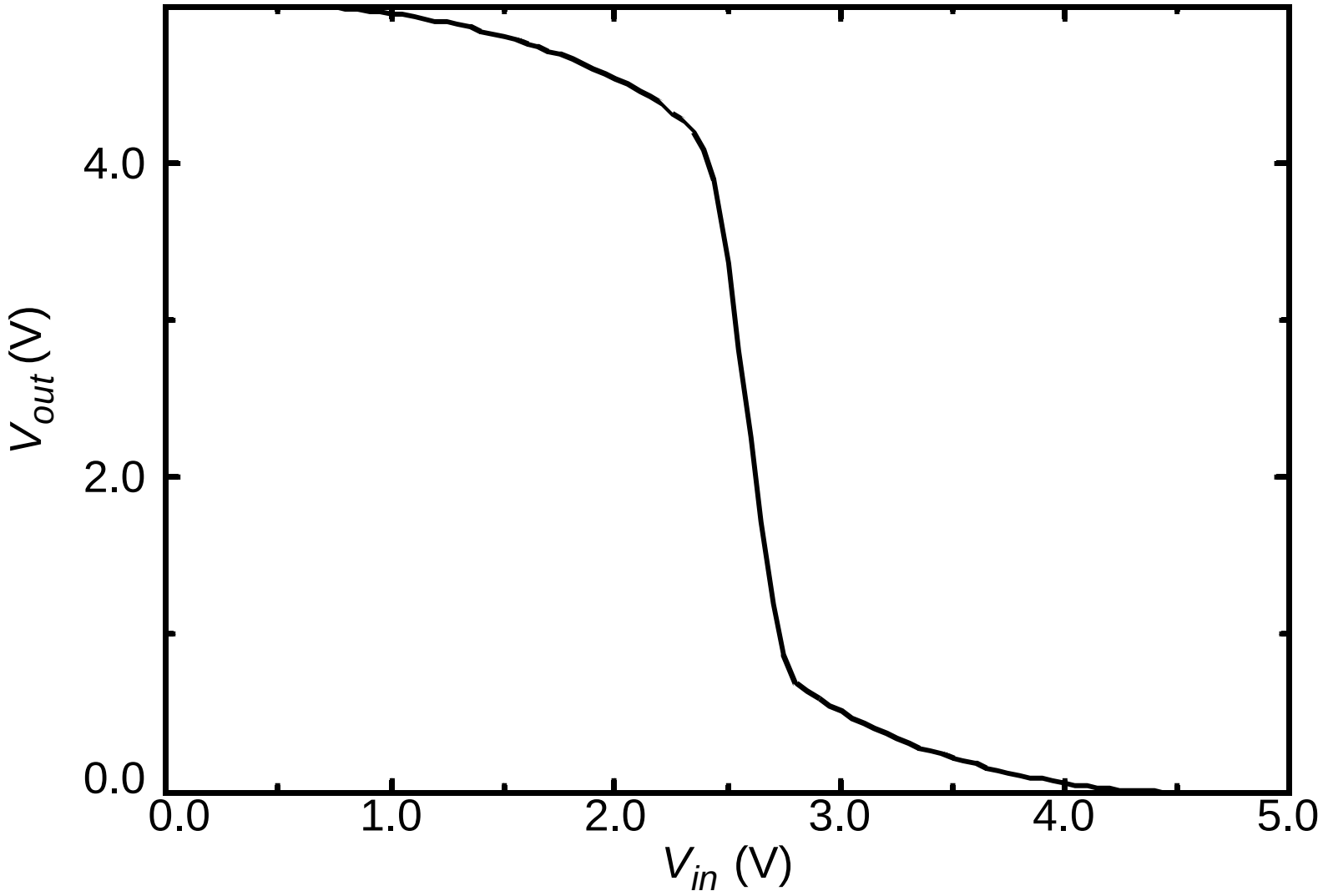
# Voltage Transfer Characteristic

# CMOS Inverter Load Characteristics

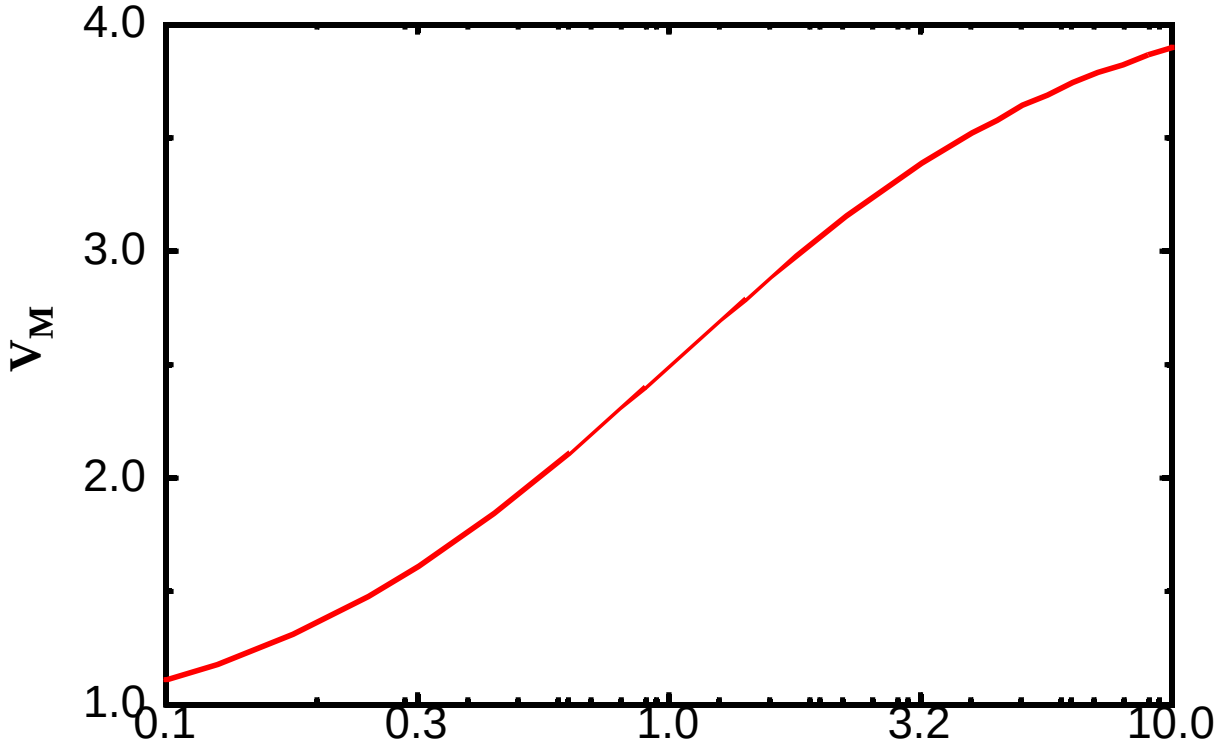




# Simulated VTC



# Gate Switching Threshold




$$V_M = \frac{r(V_{DD} - |V_{Tp}|) + V_{Tn}}{1 + r} \quad \text{with} \quad r = \sqrt{\frac{k_p}{k_n}}$$

# CMOS Inverter VTC

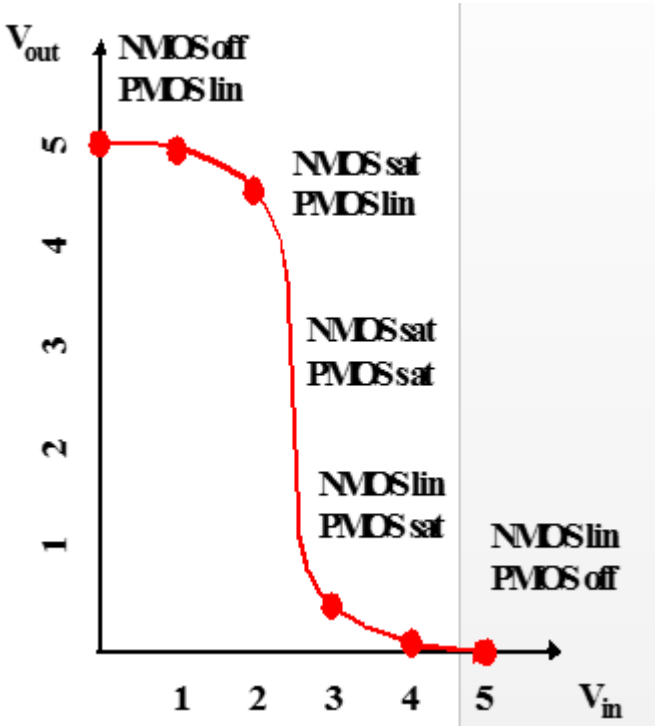
- **In linear region :**
  - Channel resistance $= R_{c_{lin}} =$

$$\lim_{V_{ds} \to 0} \left(\frac{dI_{ds}}{dV_{ds}}\right)^{-1}$$

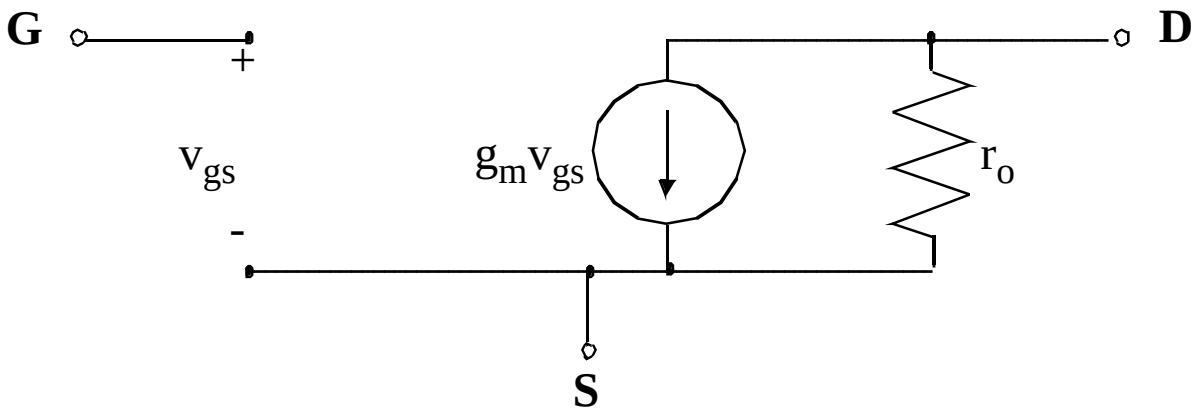
$$= \frac{1}{\beta(V_{gs} - V_T)}$$

  - Depends on $V_{gs}$
  - Transconductance $= g_m = \left(\frac{dI_{ds}}{dV_{gs}}\right) = \beta V_{ds}$
  - Higher current gain with higher $V_{ds}$

- **In saturation region :**
  - Transconductance $= g_m = \left(\frac{dI_{ds}}{dV_{gs}}\right) = \beta(V_{gs} - V_T$



# MOS Transistor Small Signal Model

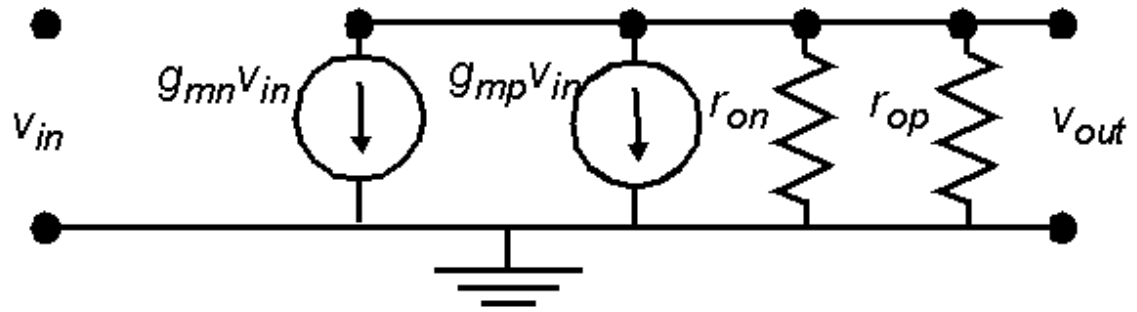



| | $g_m$ | $r_o$ |
|---|---|---|
| linear | $kV_{DS}$ | $[k(V_{GS}-V_T-V_{DS})]^{-1}$ |
| saturation | $k(V_{GS}-V_T)$ | $1/\lambda I_D$ |

# Determining $V_{IH}$ and $V_{IL}$

**At $V_{IH}$ ($V_{IL}$):** $\dfrac{\partial V_{out}}{\partial V_{in}} = -1$

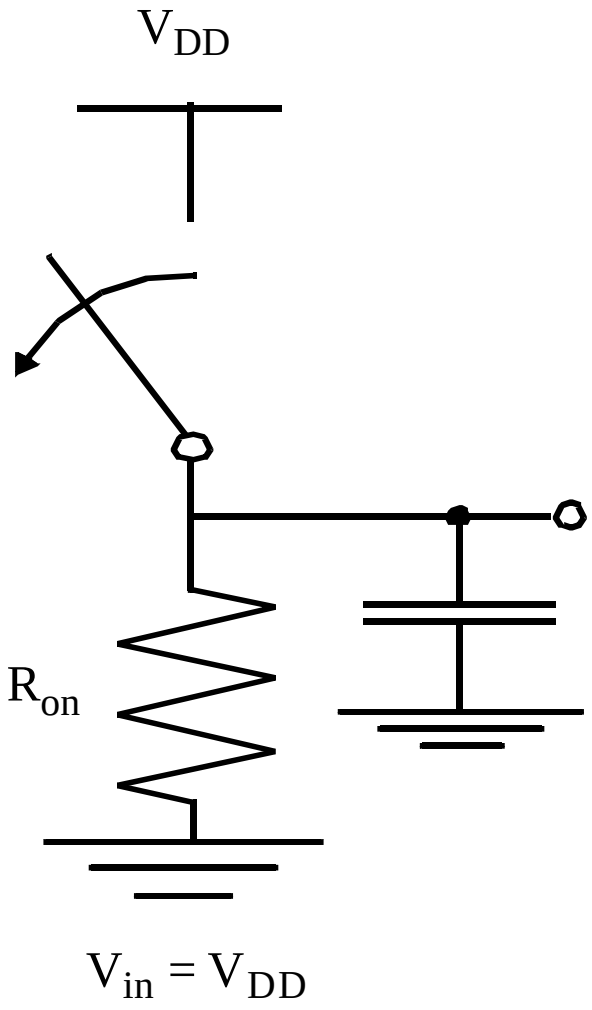
***small-signal model of inverter***




$$g = \frac{v_{out}}{v_{in}} = -(g_{mn} + g_{mp}) \times (r_{on} \| r_{op}) = -1$$

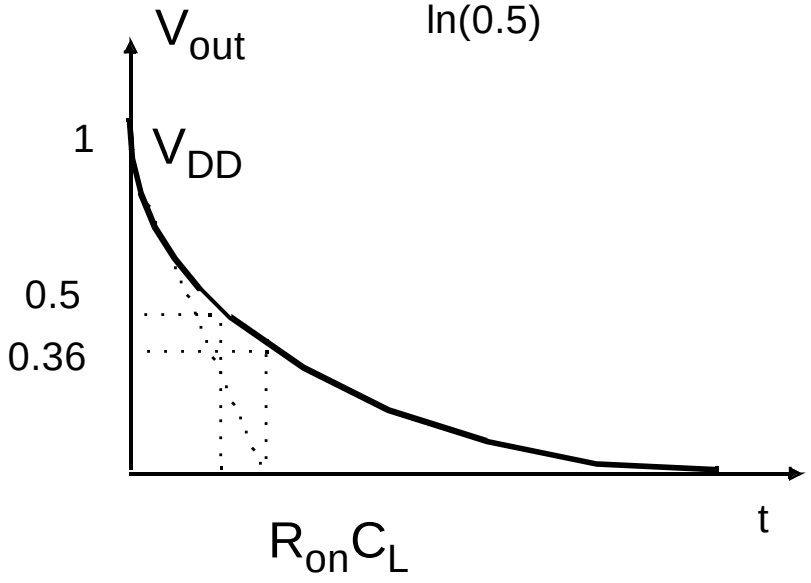
# Propagation Delay

# CMOS Inverter: Transient Response

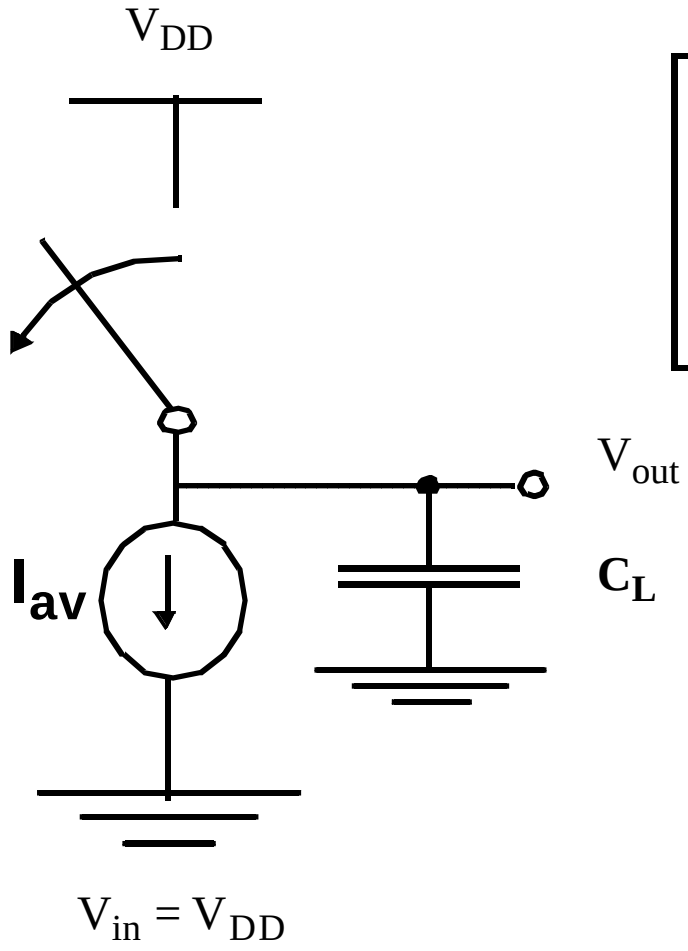



$$t_{pHL} = f(R_{on}.C_L)$$
$$= 0.69\ R_{on}C_L$$

ln(0.5)



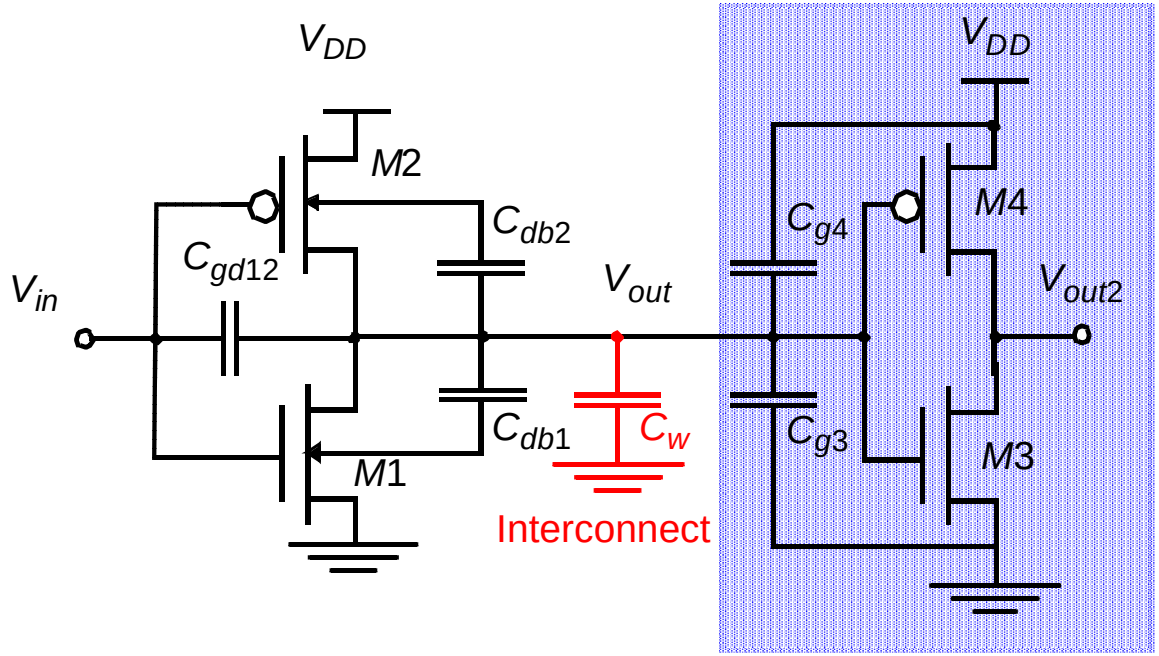
# CMOS Inverter Propagation Delay




$$t_{pHL} = \frac{C_L \, V_{swing}/2}{I_{av}}$$

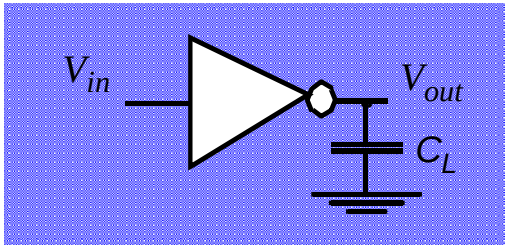
$$\sim \frac{C_L}{k_n \, V_{DD}}$$

# Computing the Capacitances



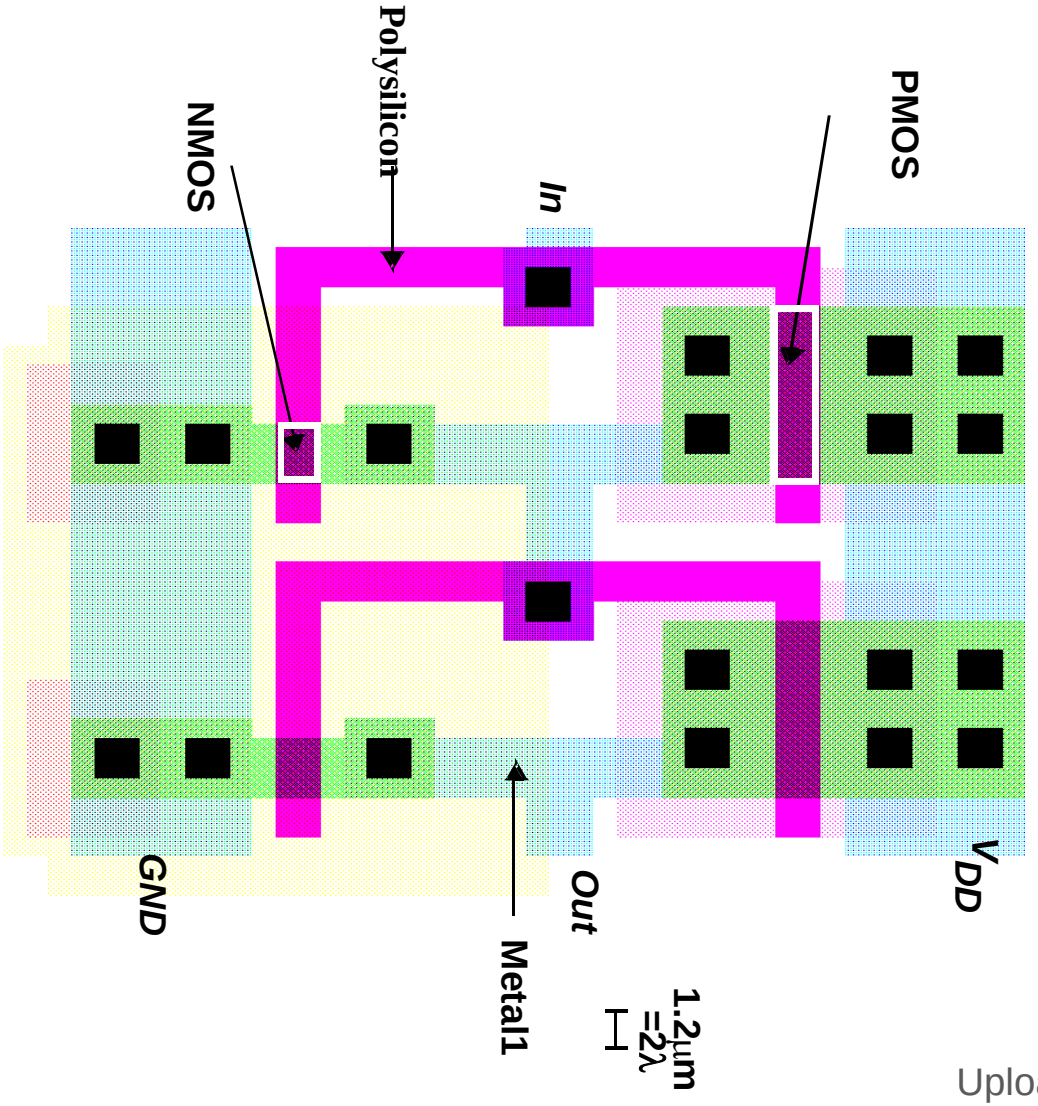

Fanout

**Simplified Model**

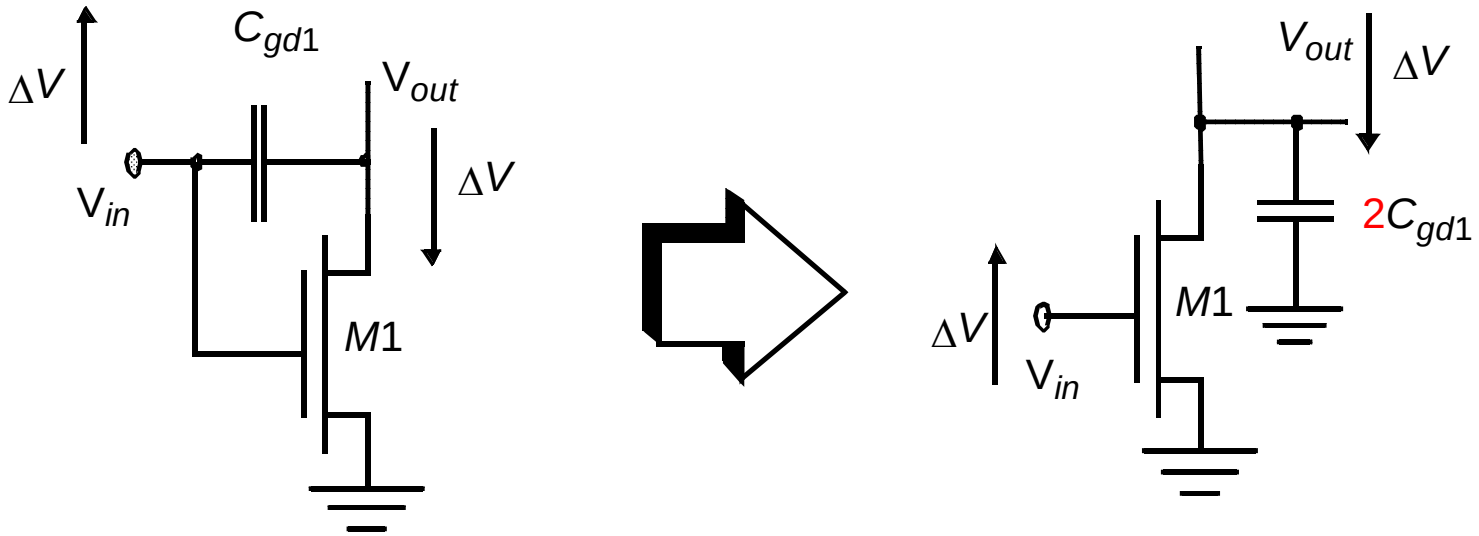



| Capacitor | Expression |
|---|---|
| $C_{gd1}$ | 2 CGD0 $W_n$ |
| $C_{gd2}$ | 2 CGD0 $W_p$ |
| $C_{db1}$ | $K_{eqn}$ ($AD_n$ CJ + $PD_n$ CJSW) |
| $C_{db2}$ | $K_{eqp}$ ($AD_p$ CJ + $PD_p$ CJSW) |
| $C_{g3}$ | $C_{ox}$ $W_n$ $L_n$ |
| $C_{g4}$ | $C_{ox}$ $W_p$ $L_p$ |
| $C_w$ | From Extraction |
| $C_L$ | |

# CMOS Inverters

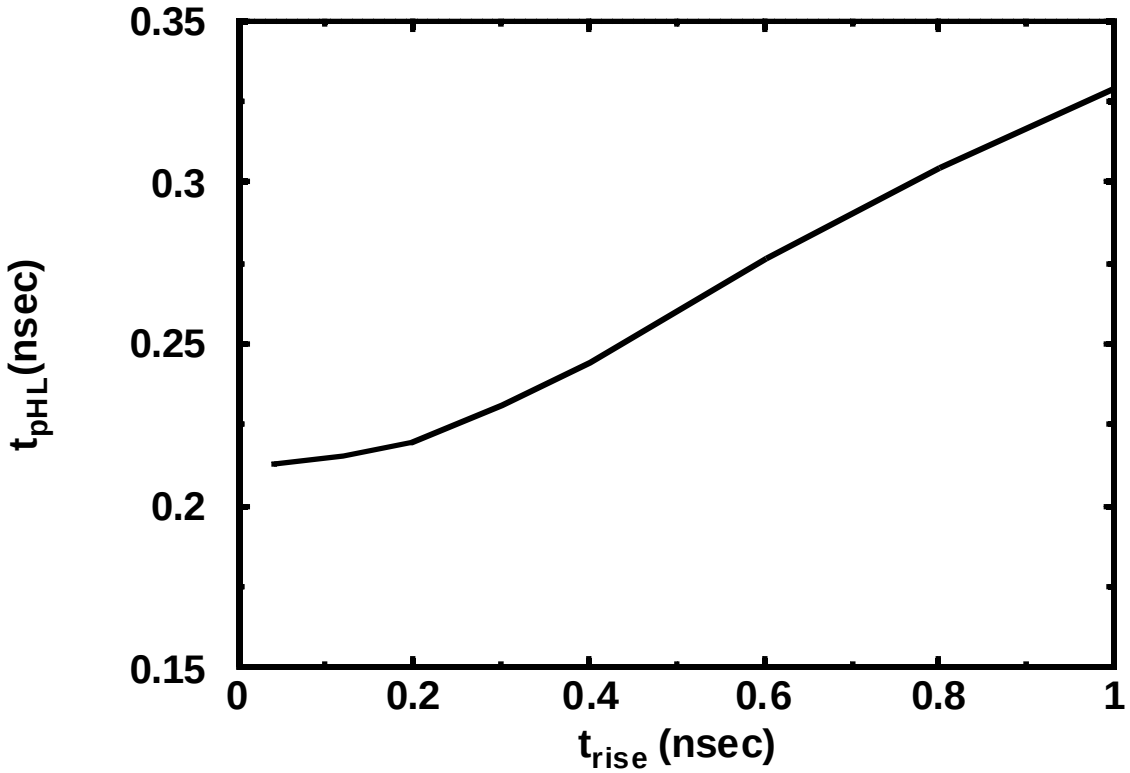


# The Miller Effect



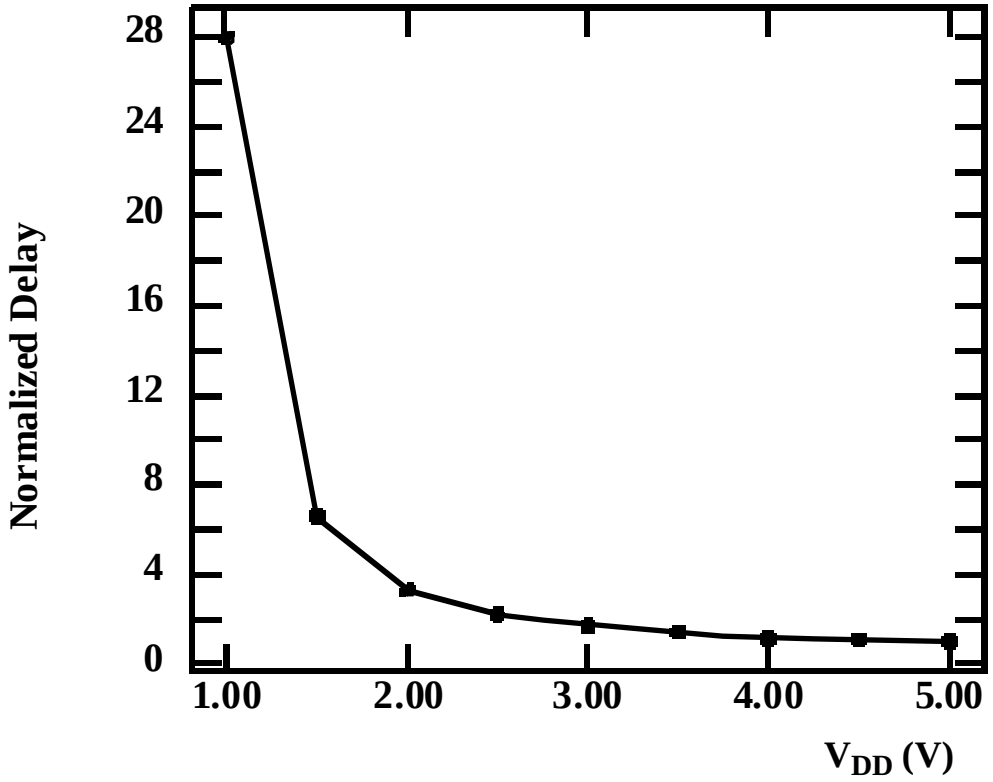

**"A capacitor experiencing identical but opposite voltage swings at both its terminals can be replaced by a capacitor to ground, whose value is two times the original value."**

# Impact of Rise Time on Delay




$$t_{pHL} = \sqrt{t_{pHL(step)}^2 + (t_r/2)^2}$$

# Delay as a function of $V_{DD}$



# Where Does Power Go in CMOS?

- **Dynamic Power Consumption**

  **Charging and Discharging Capacitors**

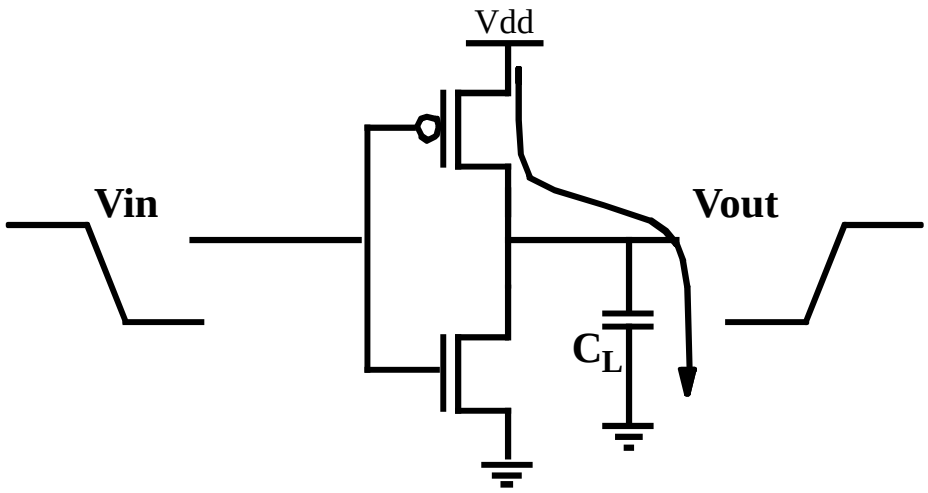
- **Short Circuit Currents**

  **Short Circuit Path between Supply Rails during Switching**

- **Leakage**

  **Leaking diodes and transistors**

# Dynamic Power Dissipation




**Energy/transition = $C_L * V_{dd}^2$**

**Power = Energy/transition * $f$ = $C_L * V_{dd}^2 * f$**

- **Not a function of transistor sizes!**
- **Need to reduce $C_L$, $V_{dd}$, and $f$ to reduce power.**

# Power Dissipation

- Energy from power supply needed to charge up the capacitor:

$$E_{charge} = \int V_{DD} i(t) dt = V_{DD} Q = V_{DD}^2 C_L$$

- Energy stored in capacitor:

$$E_{store} = 1/2 C_L V_{DD}^2$$

- Energy lost in p-channel MOSFET during charging:

$$E_{diss} = E_{charge} - E_{store} = 1/2 C_L V_{DD}^2$$

- During discharge the n-channel MOSFET dissipates an identical amount of energy. •If the charge/discharge cycle is repeated f times/second, where f is the clock frequency, the dynamic power dissipation is:

$$P = 2 E_{diss} * f = C_L V_{DD}^2 f$$

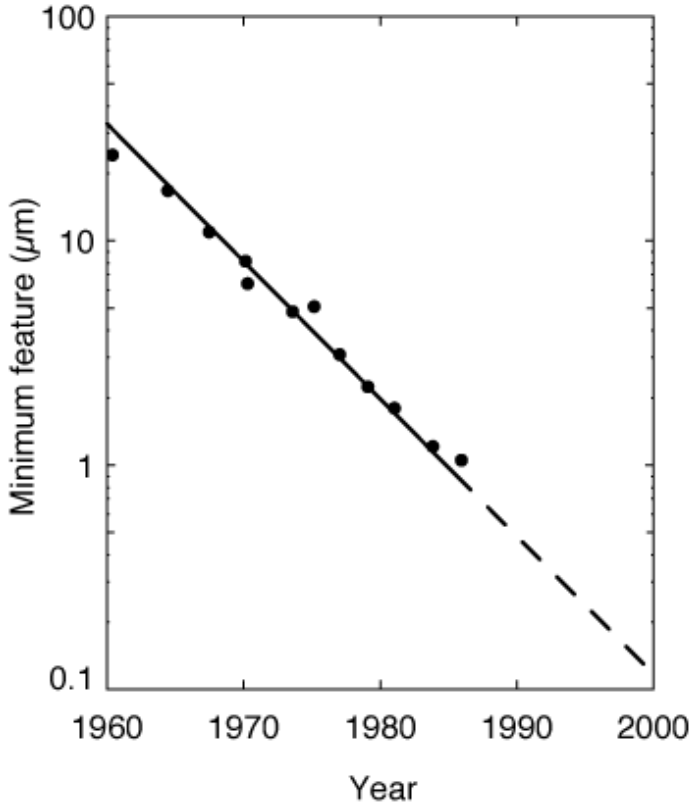
In practice many gates do not change state every clock cycle which lowers the power dissipation.

# Impact of Technology Scaling

# Technology Evolution

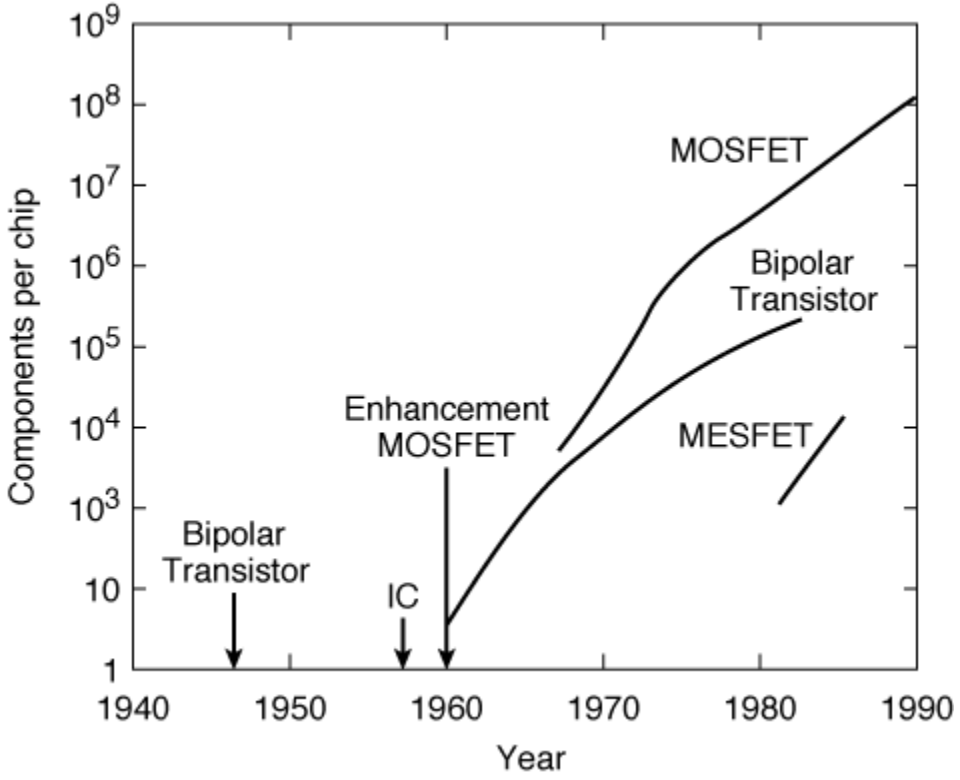
| Year of Introduction | 1994 | 1997 | 2000 | 2003 | 2006 | 2009 |
|---|---|---|---|---|---|---|
| Channel length (μm) | 0.4 | 0.3 | 0.25 | 0.18 | 0.13 | 0.1 |
| Gate oxide (nm) | 12 | 7 | 6 | 4.5 | 4 | 4 |
| $V_{DD}$ (V) | 3.3 | 2.2 | 2.2 | 1.5 | 1.5 | 1.5 |
| $V_T$ (V) | 0.7 | 0.7 | 0.7 | 0.6 | 0.6 | 0.6 |
| NMOS $I_{Dsat}$ (mA/μm) (@ $V_{GS} = V_{DD}$) | 0.35 | 0.27 | 0.31 | 0.21 | 0.29 | 0.33 |
| PMOS $I_{Dsat}$ (mA/μm) (@ $V_{GS} = V_{DD}$) | 0.16 | 0.11 | 0.14 | 0.09 | 0.13 | 0.16 |

# Technology Scaling (1)



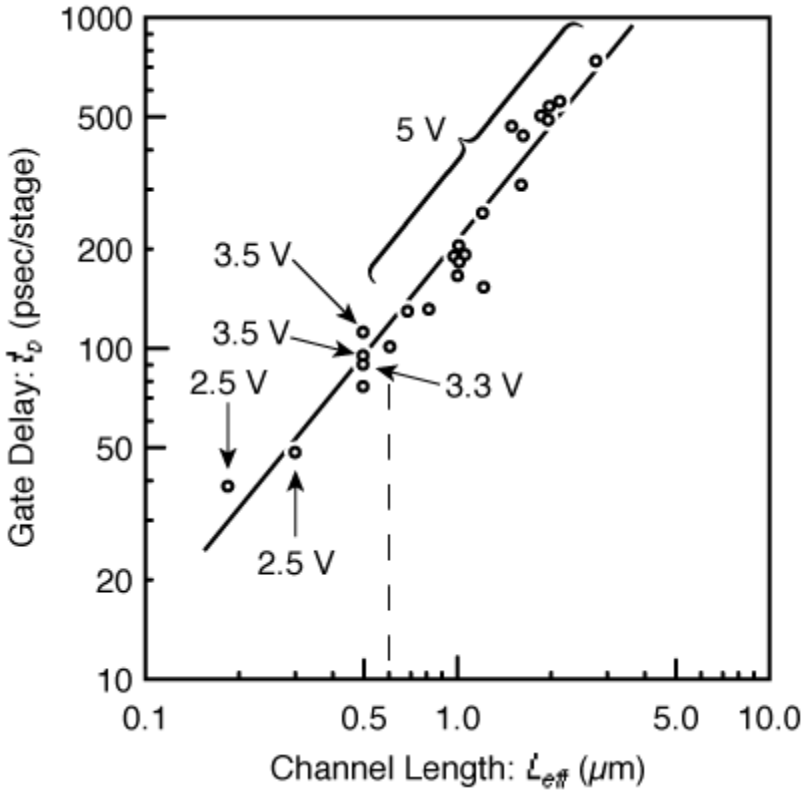

Minimum Feature Size

# Technology Scaling (2)




Number of components per chip

# Propagation Delay Scaling



# Technology Scaling Models

- **Full Scaling (Constant Electrical Field)**

  **ideal model — dimensions and voltage scale together by the same factor *S***

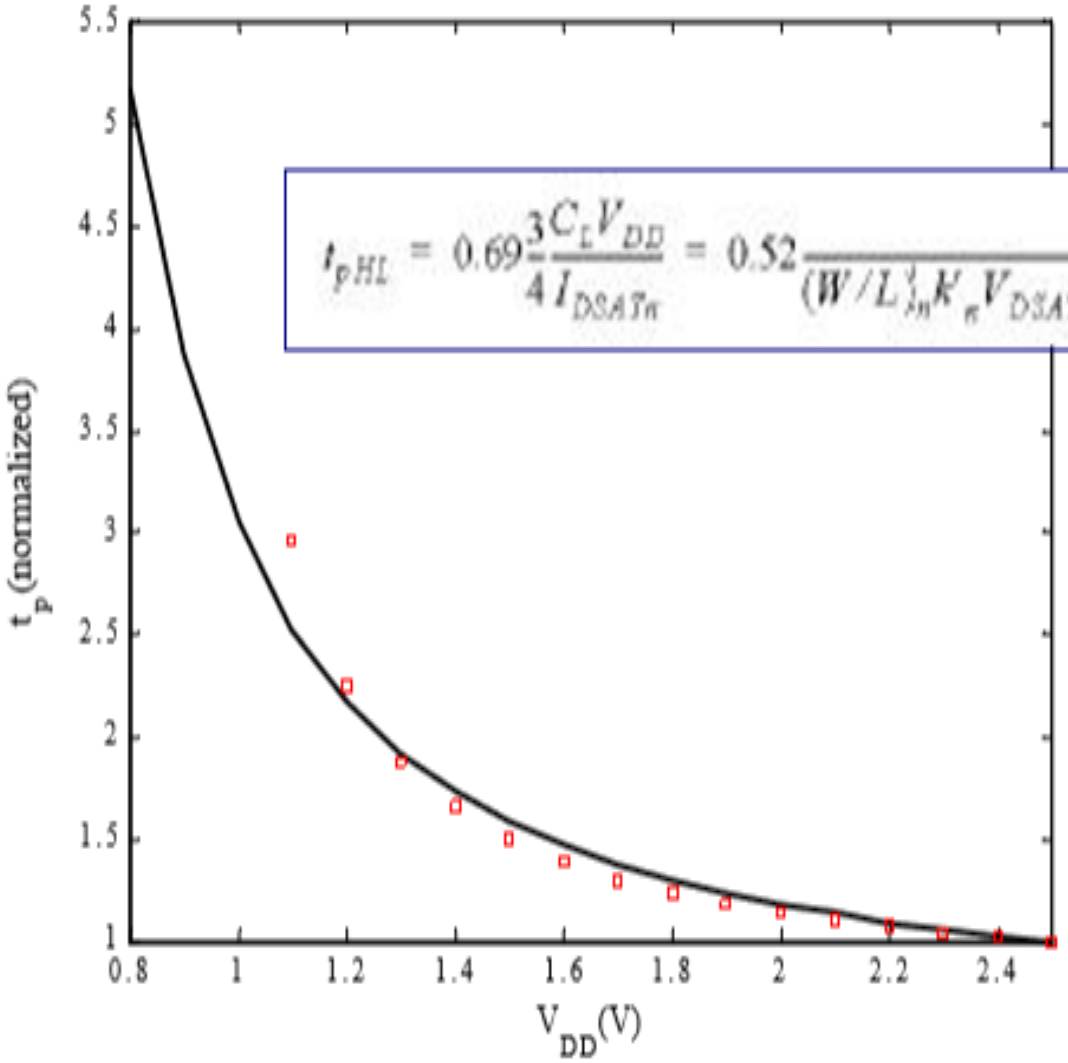
- **Fixed Voltage Scaling**

  **most common model until recently — only dimensions scale, voltages remain constant**

- **General Scaling**

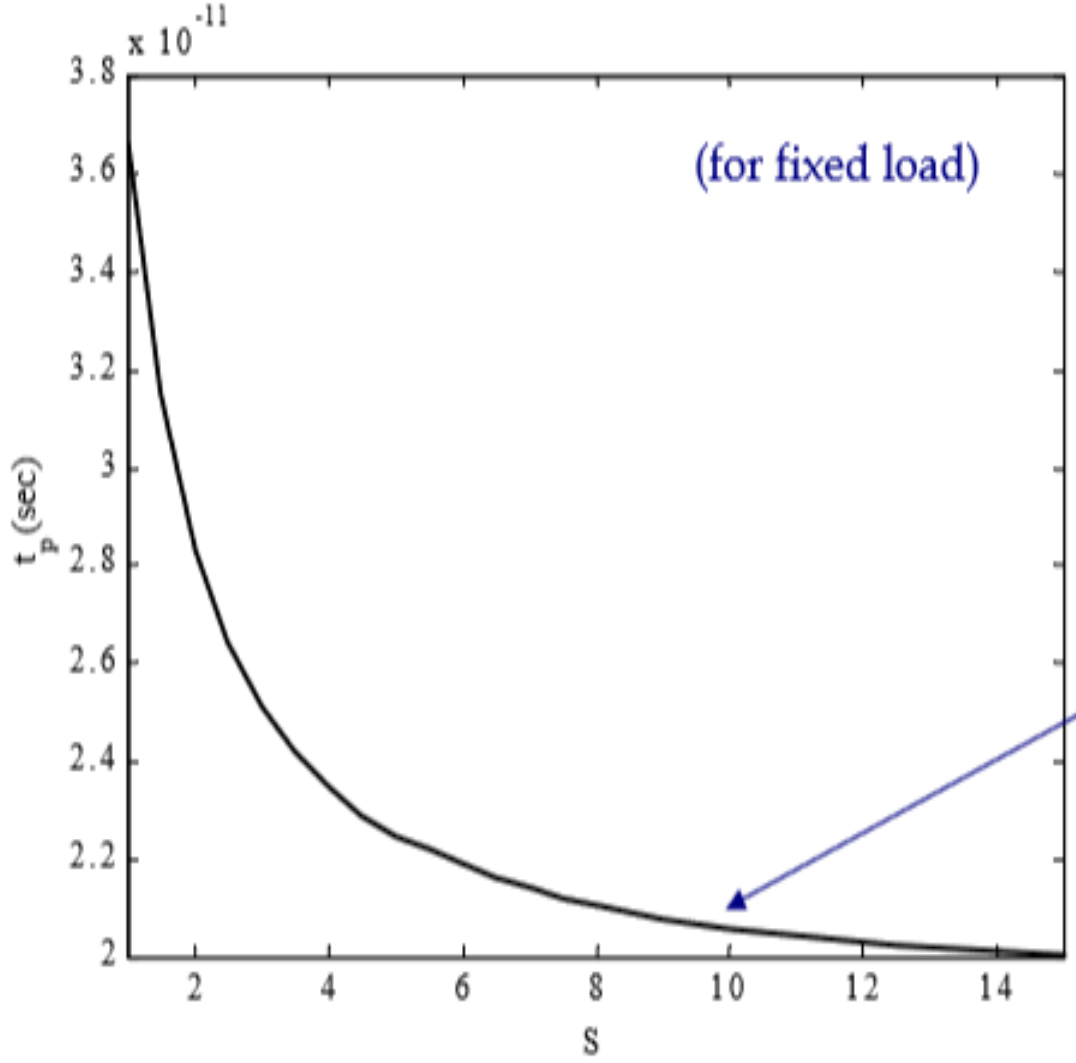
  **most realistic for todays situation — voltages and dimensions scale with different factors**

# Delay as a function of VDD



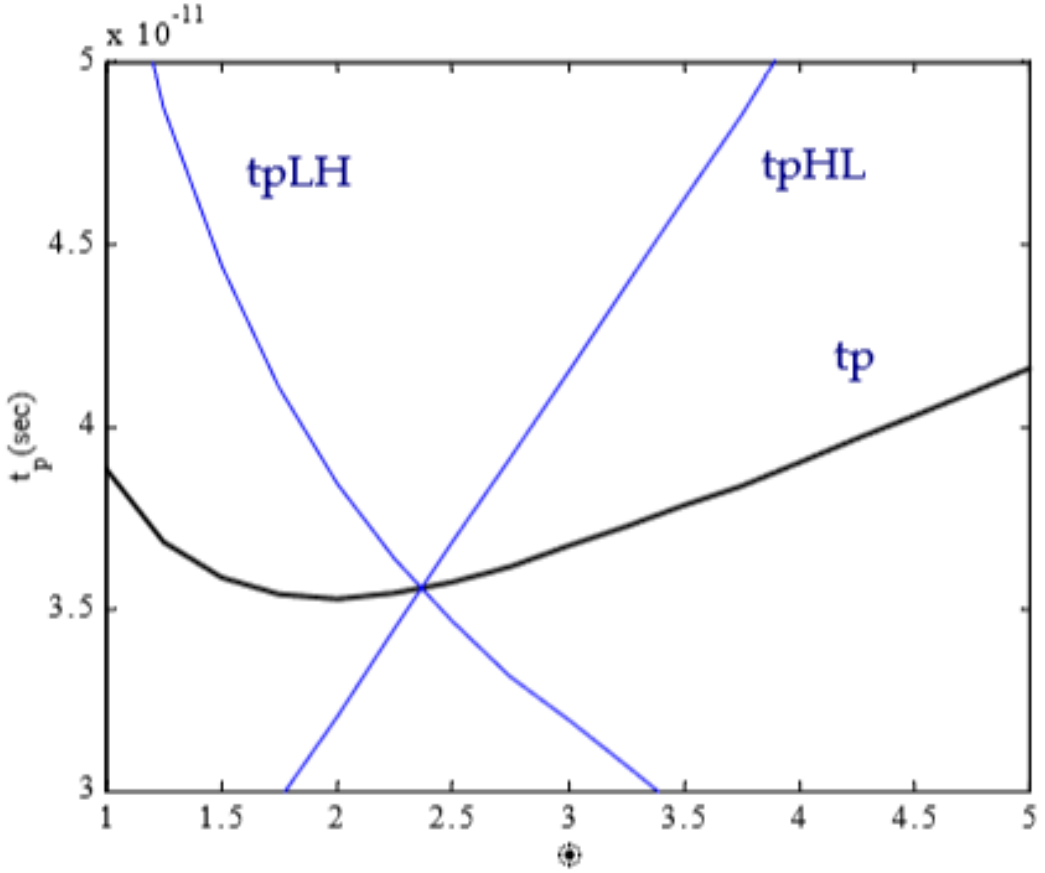

$$t_{pHL} = 0.69\frac{3C_L V_{DD}}{4 I_{DSATn}} = 0.52\frac{C_L V_{DD}}{(W/L)_n k'_n V_{DSATn}(V_{DD} - V_{Tn} - V_{DSATn}/2)}$$

# Device Sizing




Self-loading effect:
Intrinsic capacitances
dominate

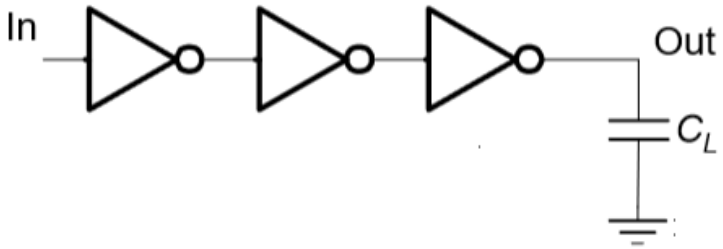
# NMOS/PMOS ratio



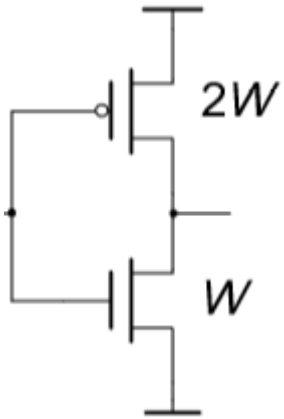

$$\beta = W_p / W_n$$

# Inverter Chain/Sizing

If CL is given: - How many stages are needed to minimize the delay?
How to size the inverters?
May need some additional constraints.




- Minimum length devices, L=0.25µm
- Assume that for WP = 2WN =2W
- same pull-up and pull-down currents
- approx. equal resistances RN = RP
- approx. equal rise tpLH and fall tpHL delays
- Analyze as an RC network




$$R_P = R_{unit}\left(\frac{W_P}{W_{unit}}\right)^{-1} \approx R_{unit}\left(\frac{W_N}{W_{unit}}\right)^{-1} = R_N = R_W$$

Delay ($D$): $t_{pHL} = (\ln 2)\, R_N C_L$ $\qquad$ $t_{pLH} = (\ln 2)\, R_P C_L$

Load for the next stage: $C_{gin} = 3\dfrac{W}{W_{unit}} C_{unit}$

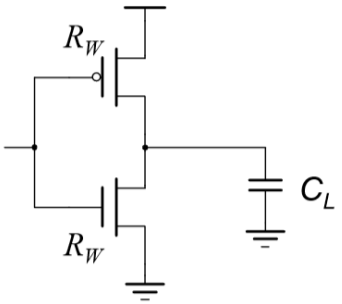
# Inverter Chain/Sizing

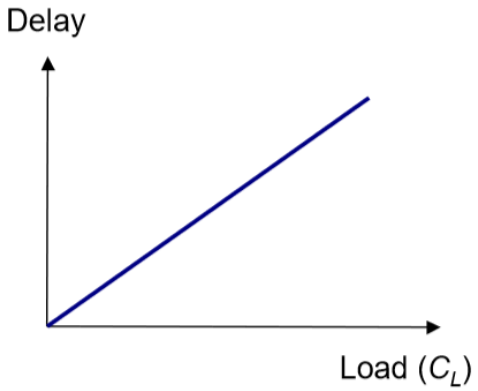
$$t_p = k R_W C_L$$

k is a constant, equal to 0.69

Assumptions: no load -> zero delay

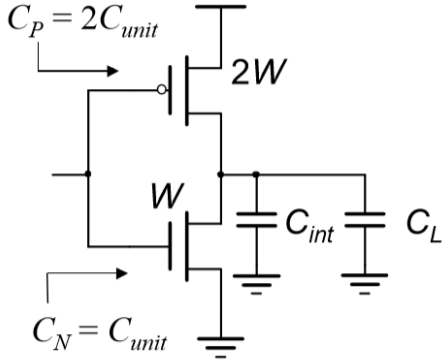
$W_{unit} = 1$

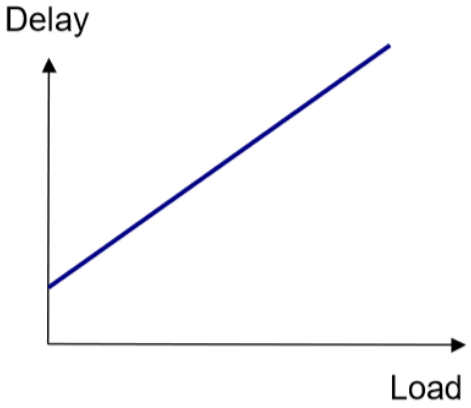



Delay = $kR_W(C_{int} + C_L) = kR_WC_{int} + kR_WC_L = kR_W C_{int}(1+ C_L /C_{int})$
= Delay (Internal) + Delay (Load)

$$Delay \sim R_W\left(C_{int} + C_L\right)$$




$$t_p = kR_W C_{\text{int}}\left(1 + C_L / C_{int}\right) = t_{p0}\left(1 + f / g\right)$$

# nFET vs. pFET

$$R_n = \frac{1}{\beta_n (V_{DD} - V_{Tn})} \qquad \beta_n = \mu_n C_{ox} \left(\frac{W}{L}\right)_n$$

$$R_p = \frac{1}{\beta_p (V_{DD} - |V_{Tp}|)} \qquad \beta_p = \mu_p C_{ox} \left(\frac{W}{L}\right)_p$$

$$\frac{\mu_n}{\mu_p} = r$$ Typically (2 .. 3)

**(μ is the carrier mobility through device)**

**(We will return to this later …)**