The Inverter

References:

Adapted from: *Digital Integrated Circuits: A Design Perspective*, J. Rabaey, Prentice Hall © UCB

Principles of CMOS VLSI Design: A Systems Perspective, N. H. E. Weste, K. Eshraghian, Addison Wesley

Regions of Operation

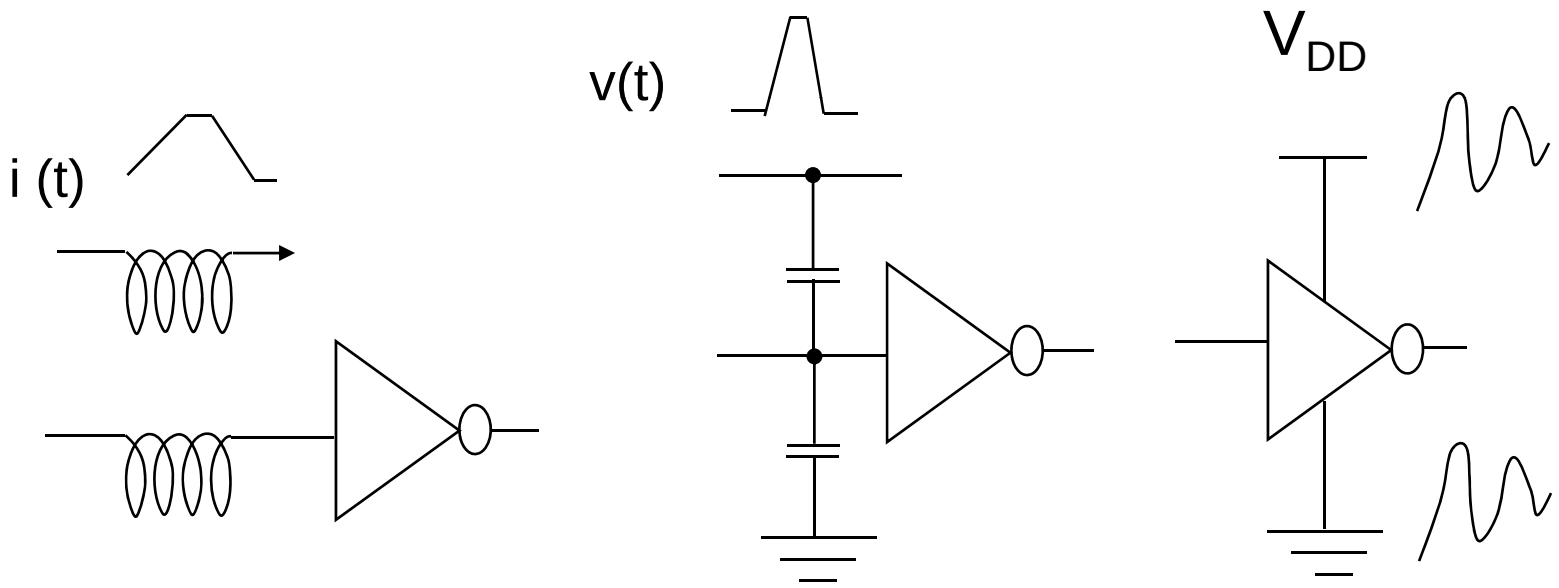
	Cutoff	Non-saturated	Saturated
p-device	$V_{gsp} > V_{tp}$ $V_{in} > V_{tp} + V_{DD}$	$V_{gsp} < V_{tp}$ $V_{in} < V_{tp} + V_{DD}$ $V_{dsp} > V_{gsp} - V_{tp}$ $V_{out} > V_{in} - V_{tp}$	$V_{gsp} = V_{tp}$ $V_{in} < V_{tp} + V_{DD}$ $V_{dsp} < V_{gsp} - V_{tp}$ $V_{out} < V_{in} - V_{tp}$
n-device	$V_{gsn} < V_{tn}$ $V_{in} < V_{tn}$	$V_{gsn} > V_{tn}$ $V_{in} > V_{tn}$ $V_{dsn} < V_{gs} - V_{tn}$ $V_{out} < V_{in} - V_{tn}$	$V_{gsn} > V_{tn}$ $V_{in} > V_{tn}$ $V_{dsn} > V_{gs} - V_{tn}$ $V_{out} > V_{in} - V_{tn}$

Digital Gates
Fundamental Parameters

- Area and Complexity
- Robustness and Reliability
- Performance
- Power Consumption

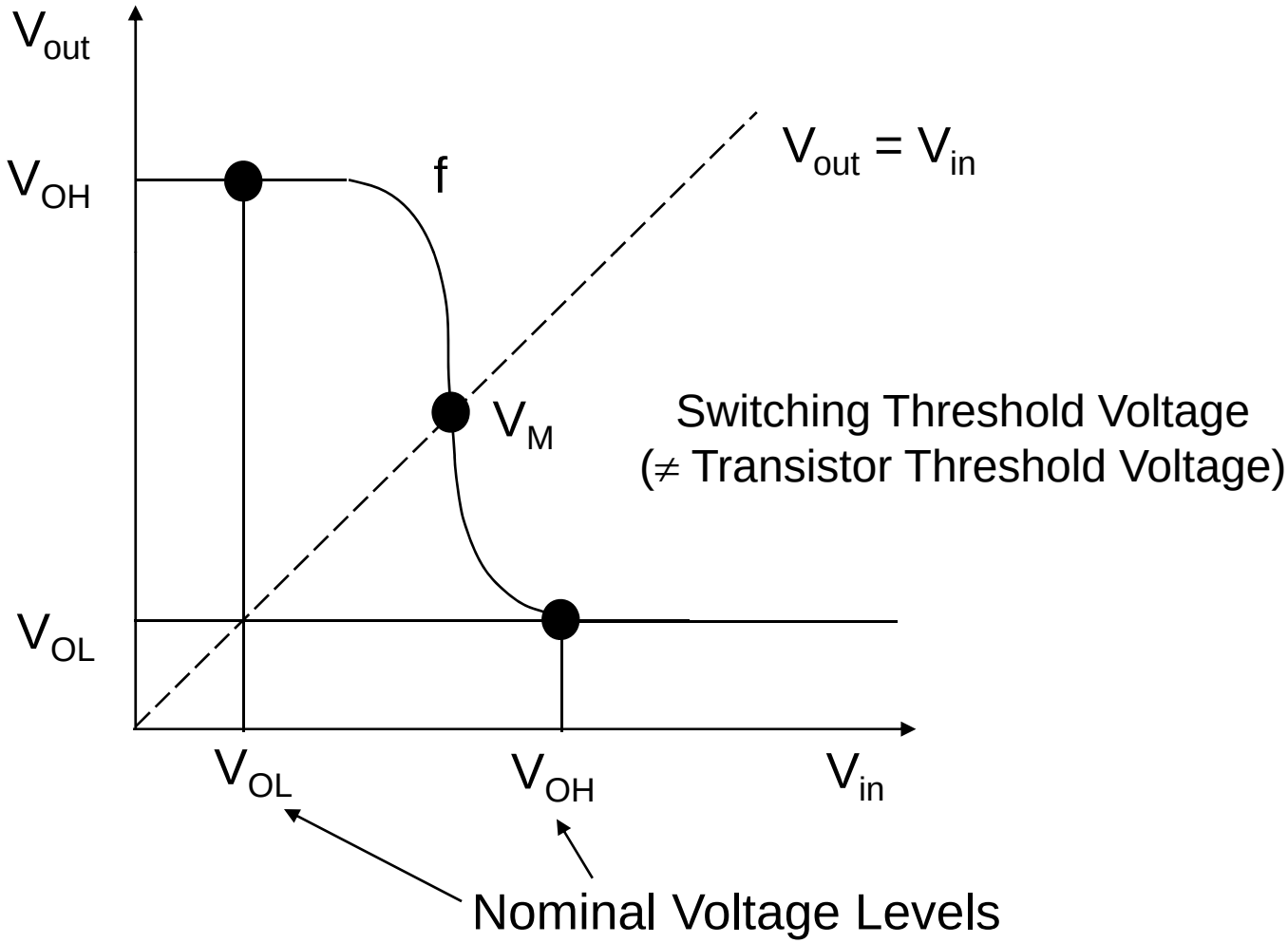
Noise in digital Integrated Circuits

unwanted variations of voltages and currents at the logic nodes

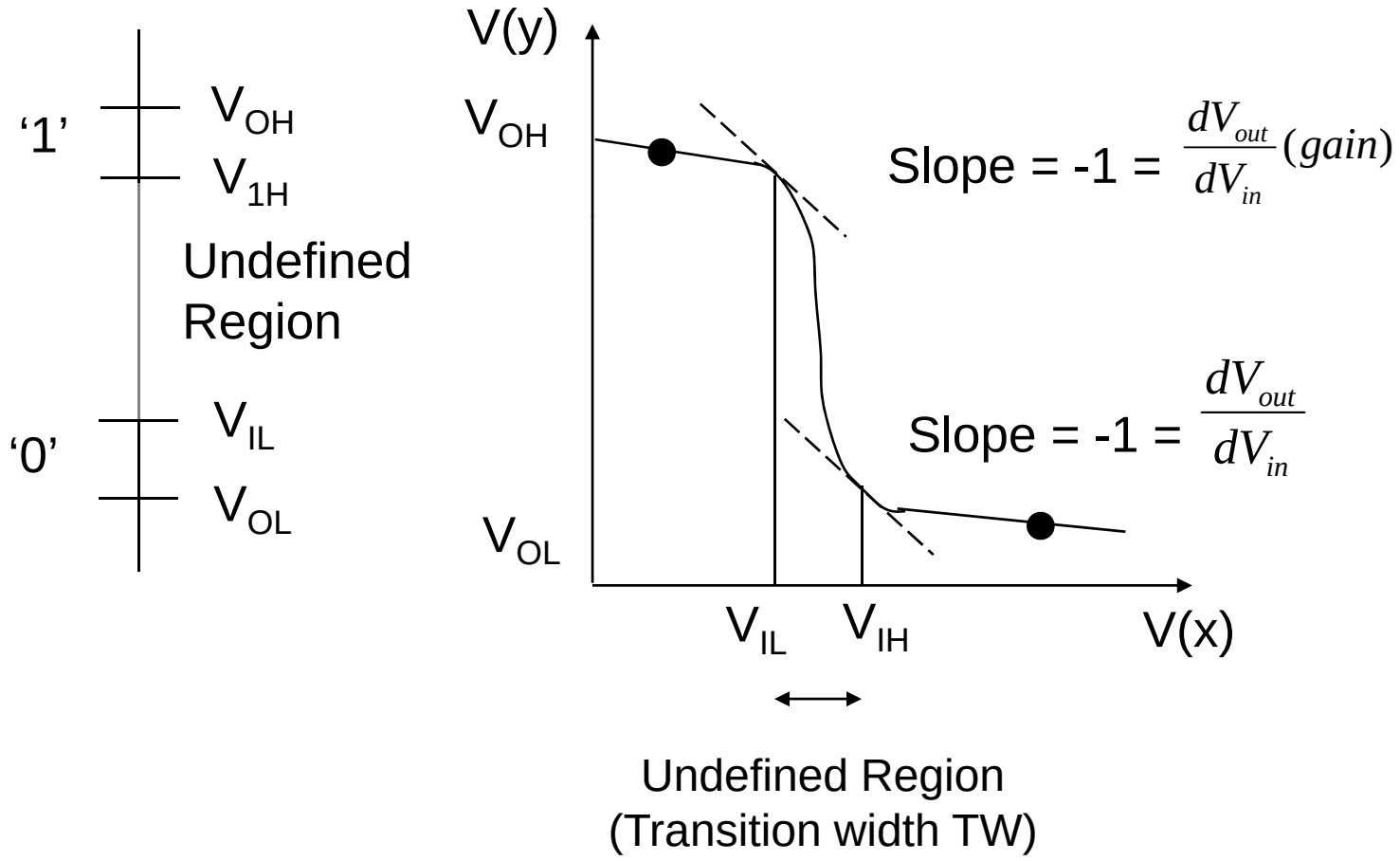


(a) Inductive coupling (b) Capacitive coupling (c) Power and ground noise

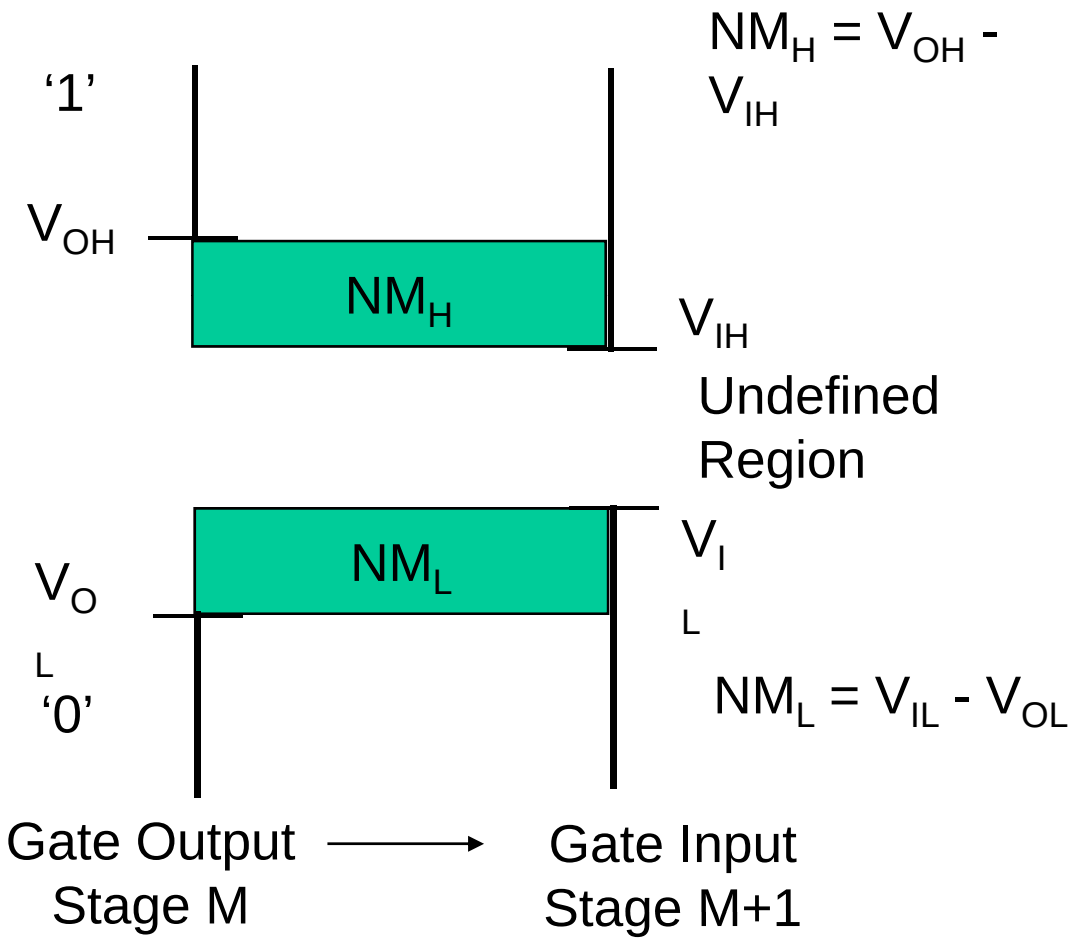
DC Operation:
Voltage Transfer Characteristic (VTC)



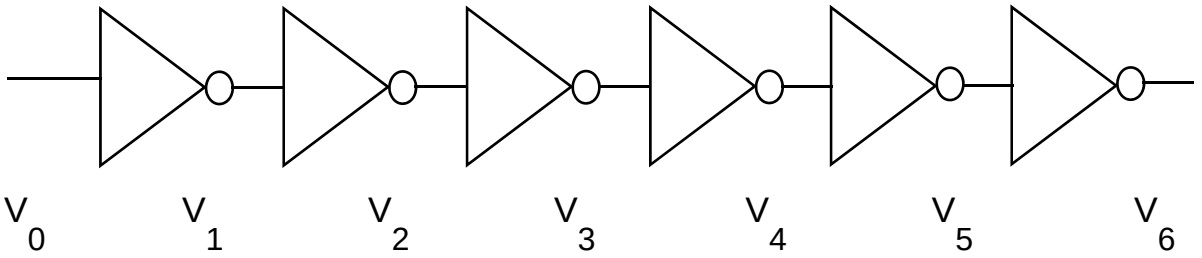
Mapping between analog and digital signals



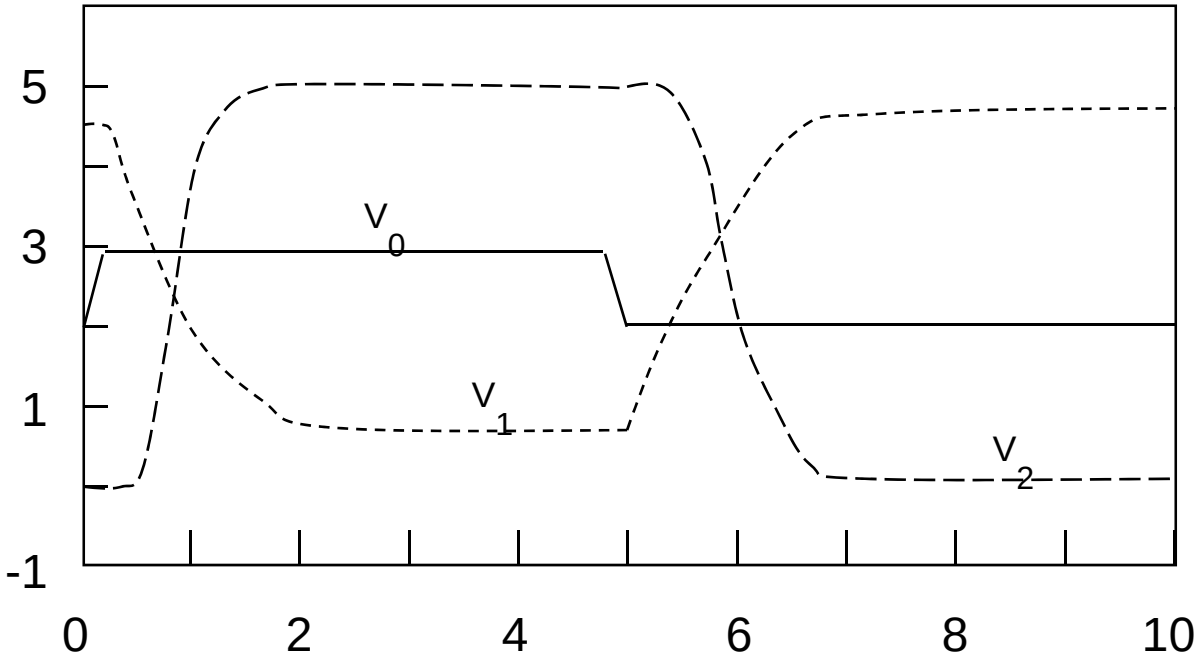
Definitaion of Noise Margins



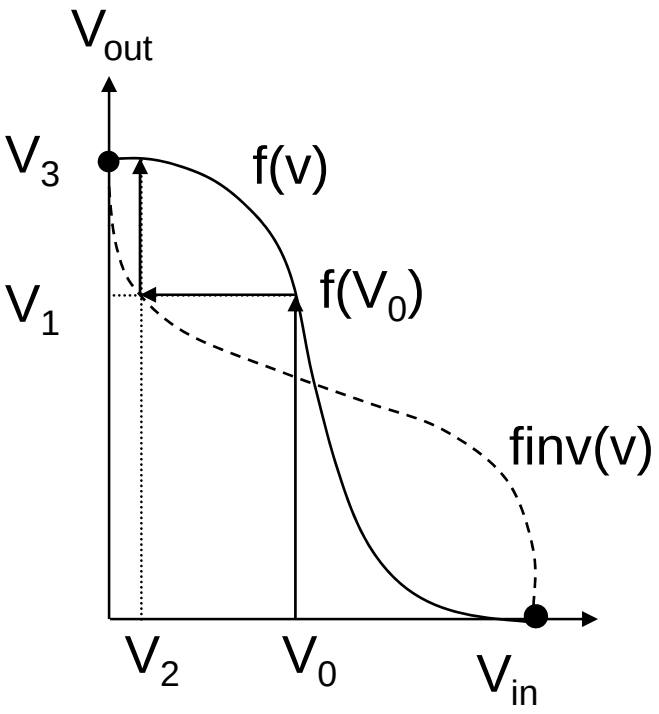
The Regenerative Property



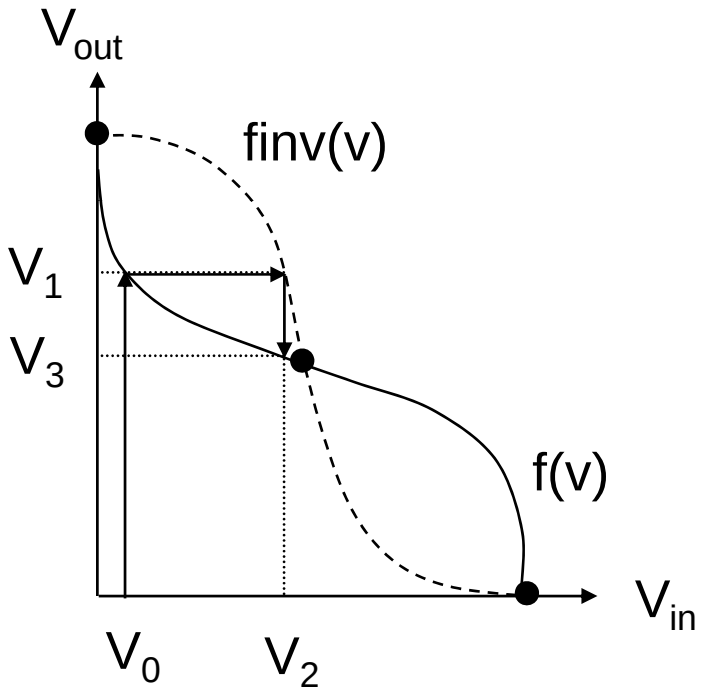
A chain of inverters



Conditions for Regeneration

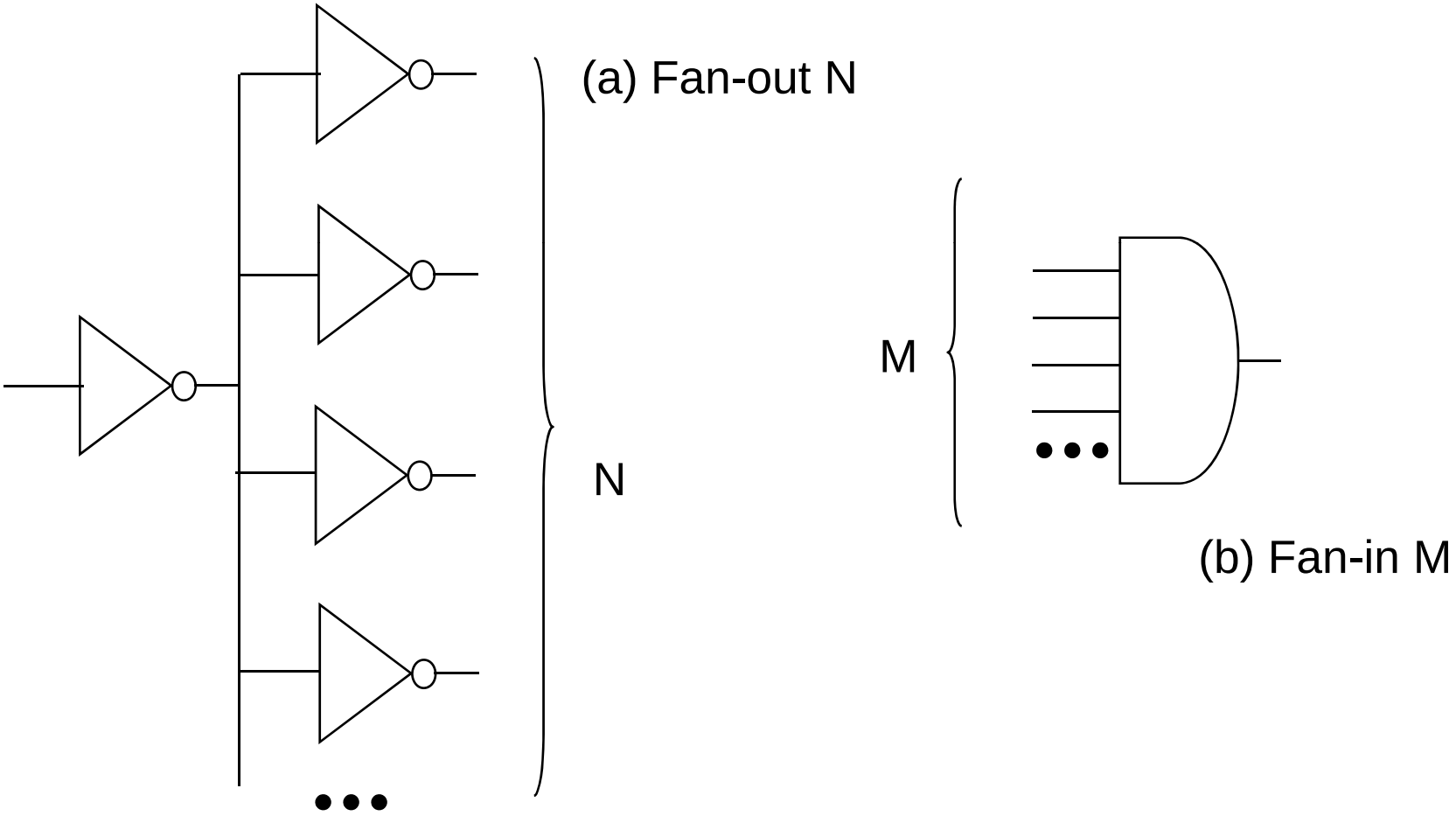


(a) Regenerative gate

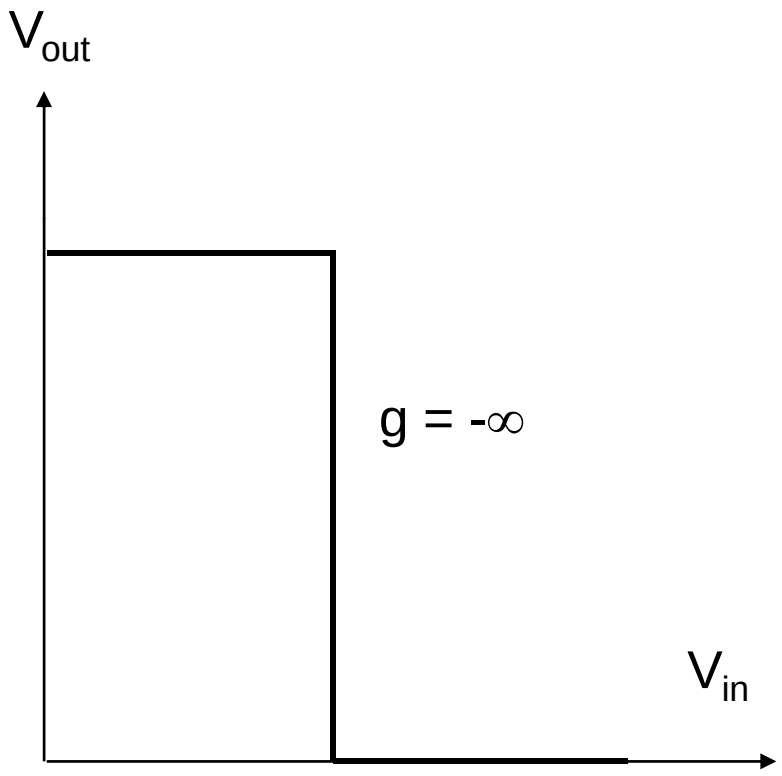


(b) Non-regenerative gate

Fan-in and Fan-out



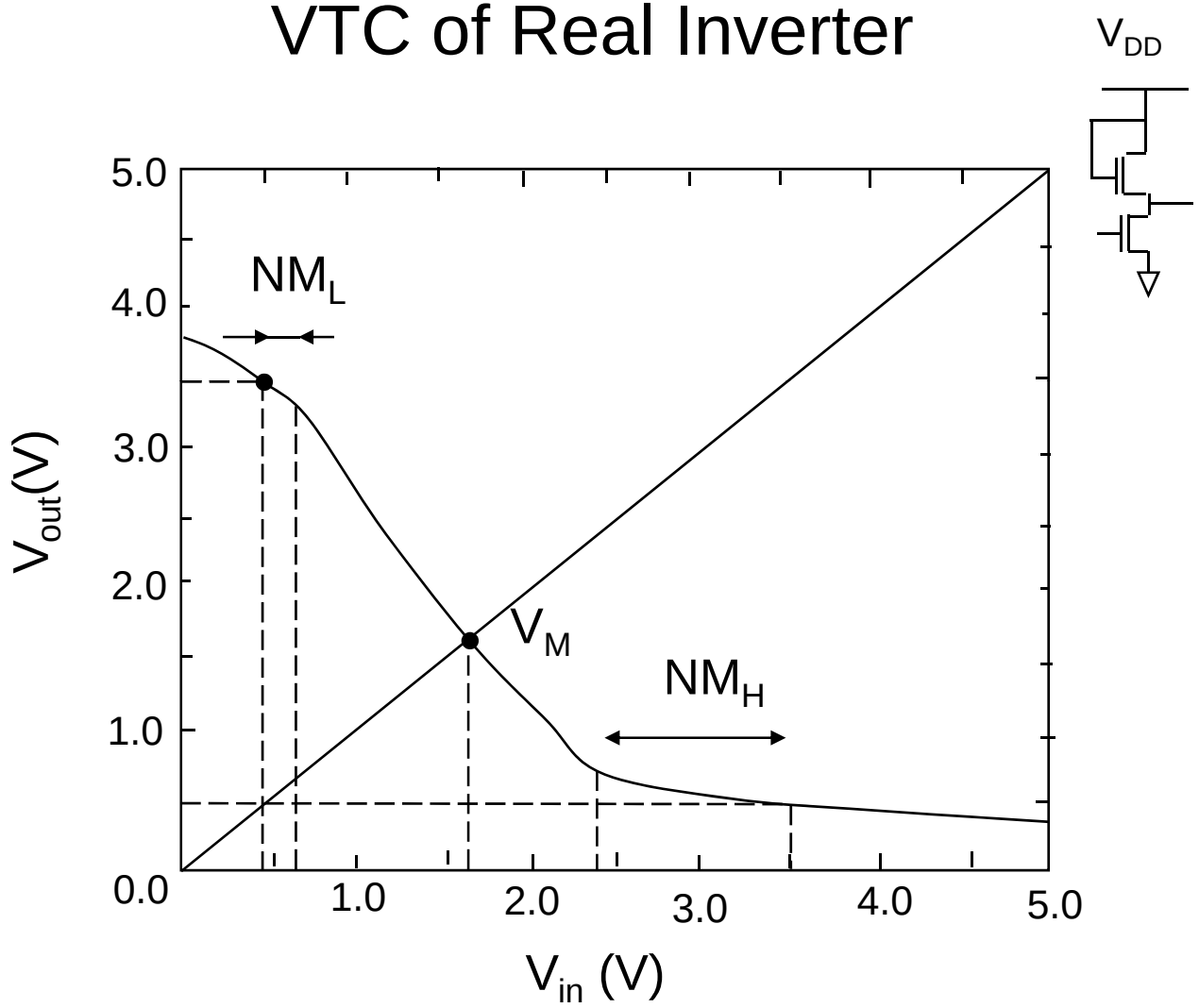
The Ideal Gate



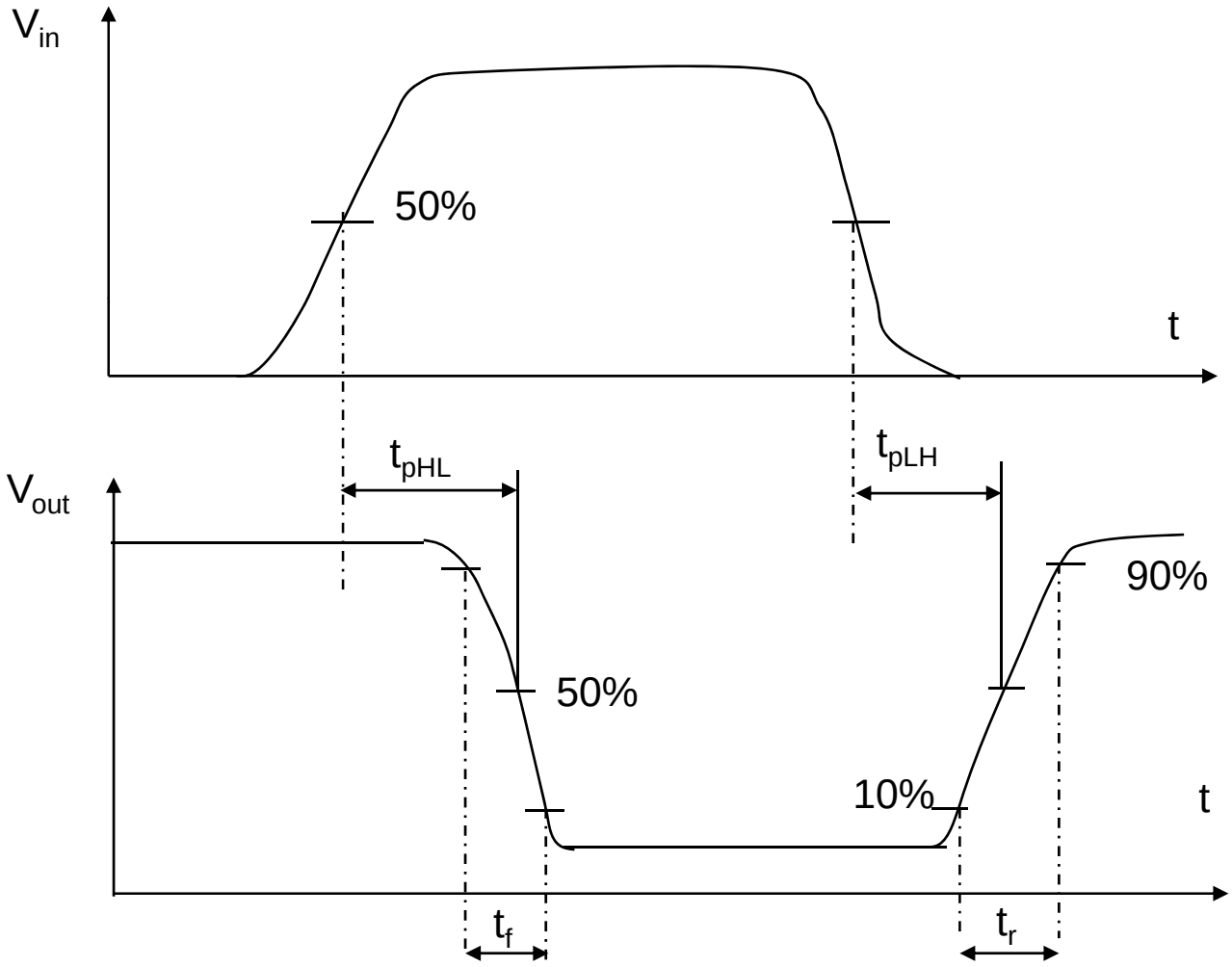
$R_i = \infty$

$R_o = 0$

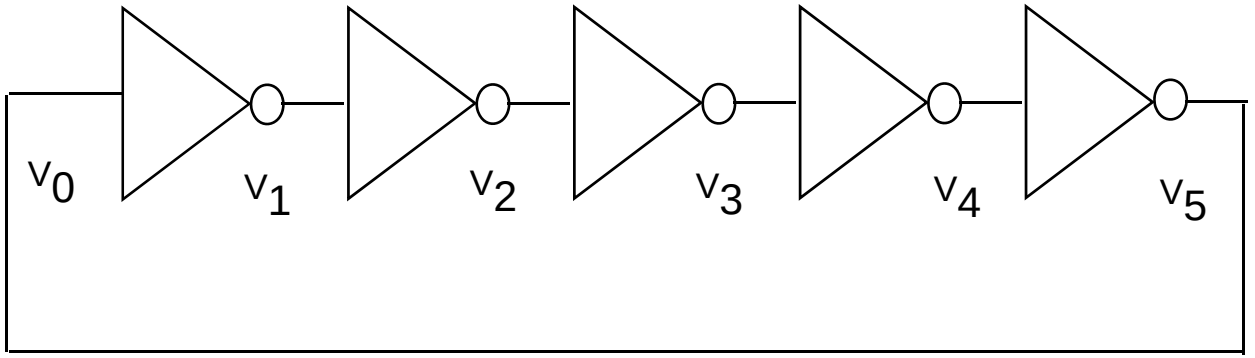
VTC of Real Inverter

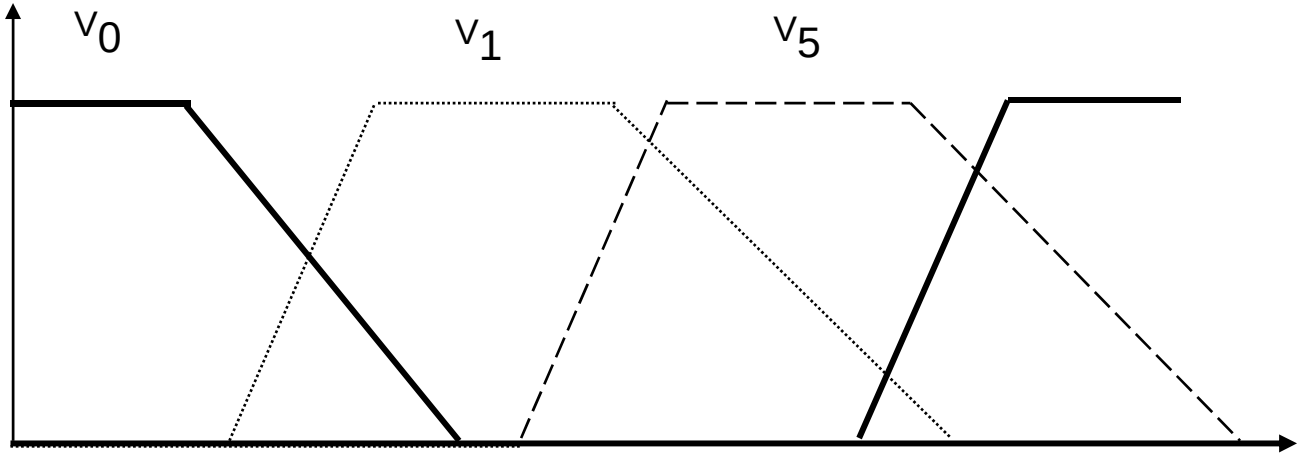


Delay Definitions



Ring Oscillator





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

$$2Nt_p >> t_f + t_r$$

Power Dissipation

P(t) = instantaneous power

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

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

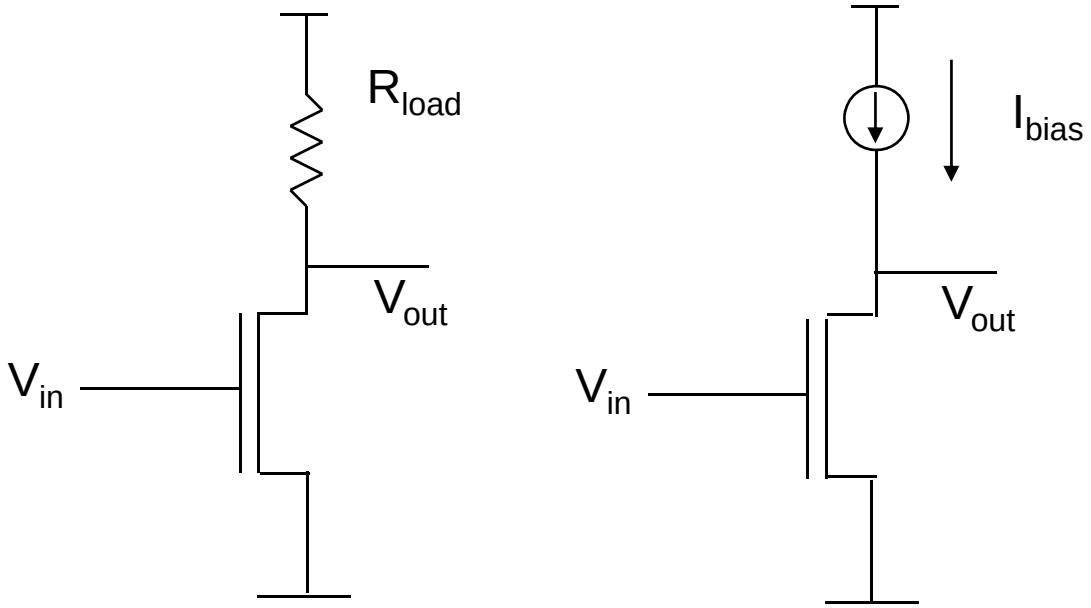
Power-Delay Product

$PDP = t_p \times P_{av}$

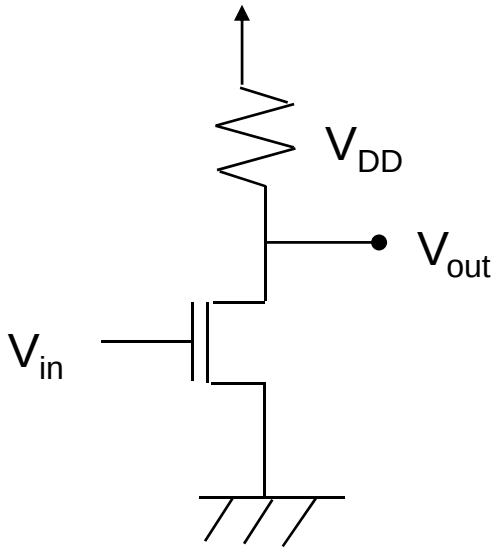
= Energy dissipated per operation

Static Load MOS Inverters

Static Load MOS Inverters

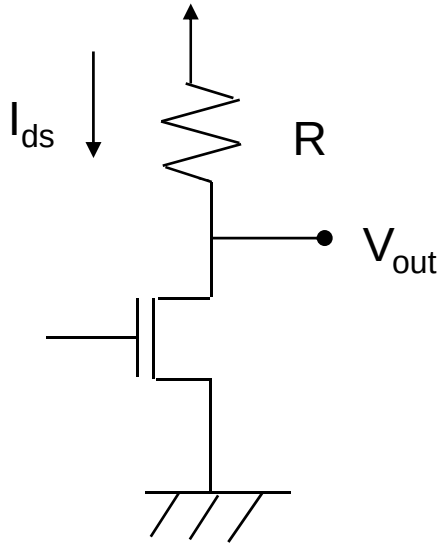


Basic Inverter



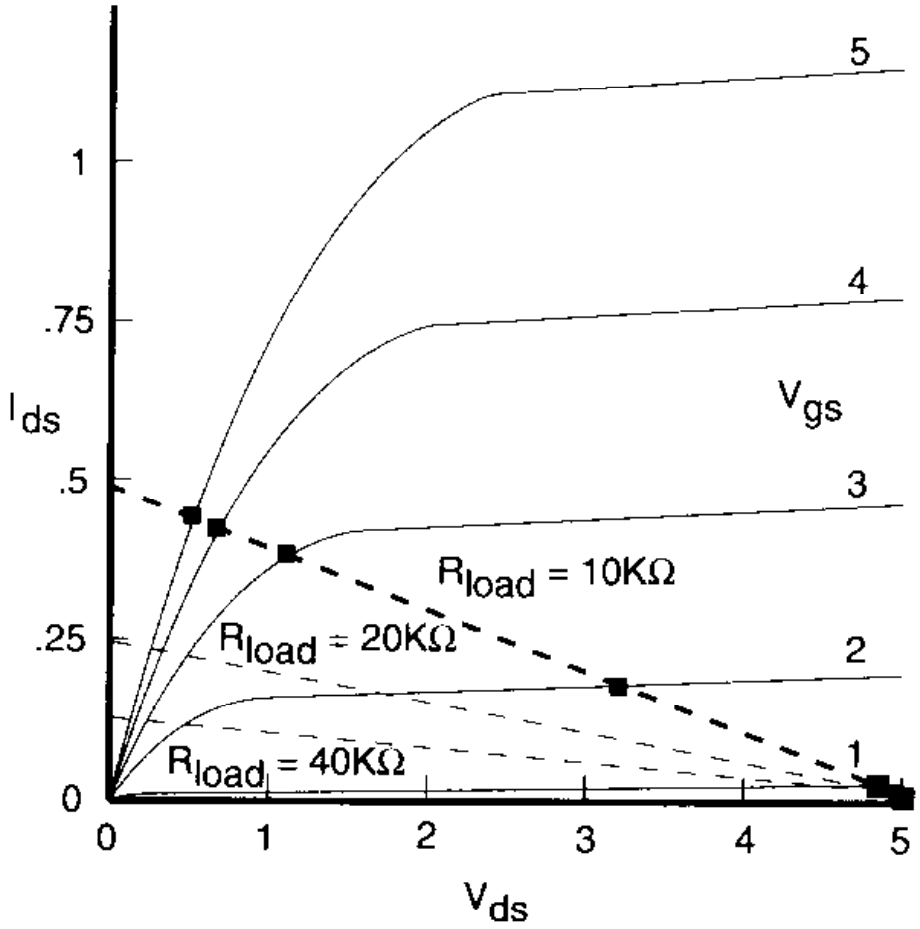
- $V_{in} < V_{th}$; NMOS off; V_{out} pulled to V_{DD}
- $V_{in} > V_{th}$; NMOS on, current flows through R to ground
- If R is sufficiently large, V_{out} could be pulled down well below V_{th};

Static Load MOS INverter



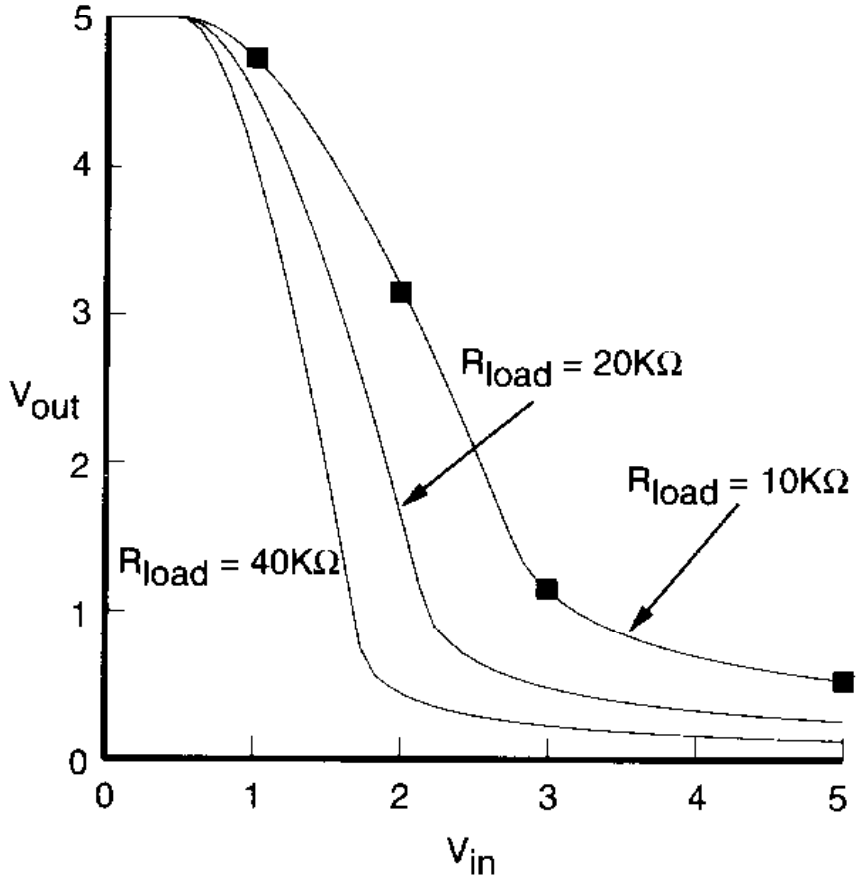
$V_{out} = V_{ds}$

$I_{ds}.R = V_{DD}-V_{ds}$



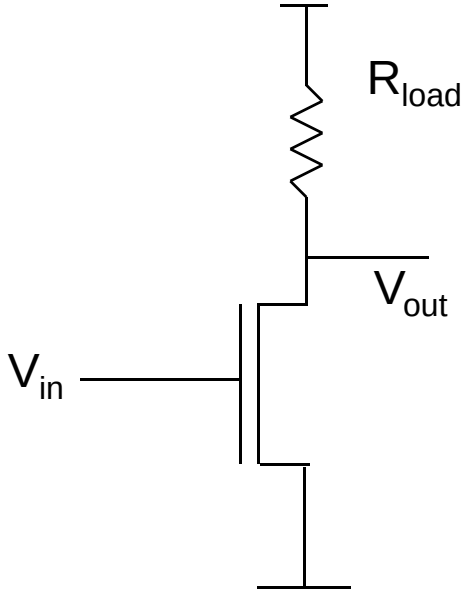
(b)

VTC of Resistive Load



(c)

Resistive Load Device



V_{oh} = 5.0V

V_{ol} = ???

$I = (V_{dd}-V_{ol})/R$

$I = \beta.((V_{dd}-V_t)V_{ol}-0.5V_{ol}^2)$

$$R = \frac{(V_{dd} - V_{ol})}{\beta.((V_{dd} - V_t)V_{ol} - 0.5V_{ol}^{\ 2})}$$

Sizing for V_{OL}

$$R = \frac{(V_{dd} - V_{ol})}{\beta.((V_{dd} - V_t)V_{ol} - 0.5V_{ol}^{\ 2})}$$

Assume: $V_{dd} = 5.0V$
$V_t = 1.0V$
$\beta = 10^{-4} A/V$

Proper design: $V_{ol} < V_t$

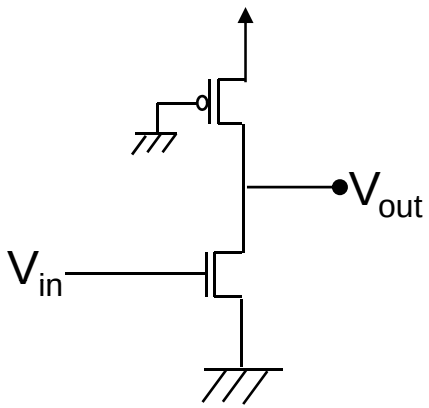
Let: $V_{ol} = 0.5V$

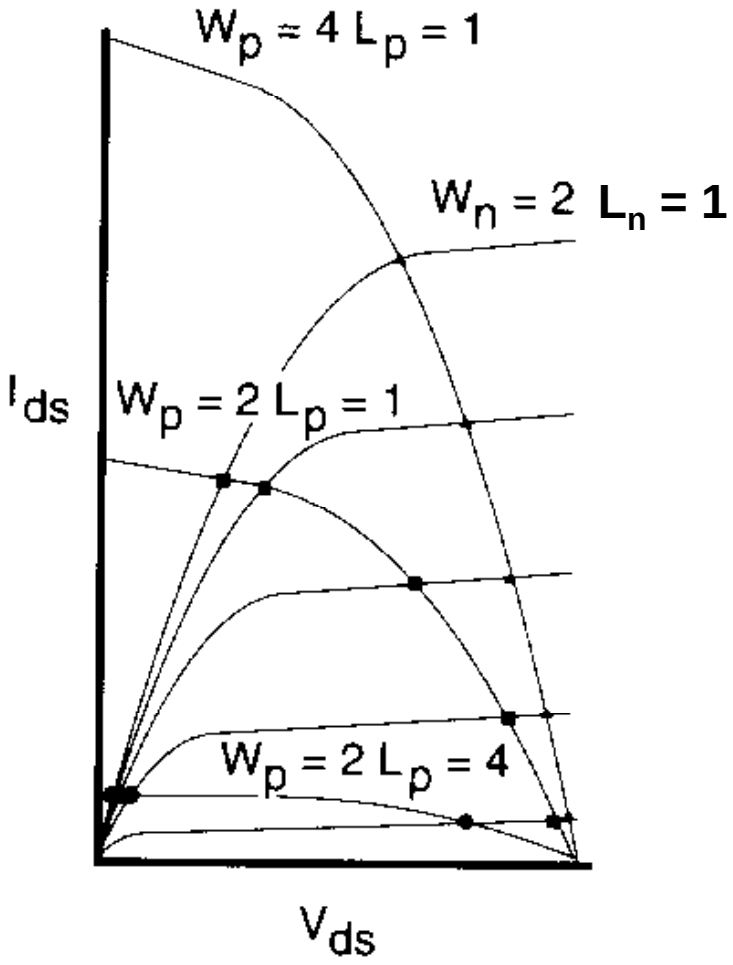
$R = 24k\Omega$

Resistor and Current-Source Loads

- Resistance/length of minimum-width lines of various connecting elements is far less than effective resistance of the switched on MOSFET
- In some memory processes, resistors are implemented by highly resistive undoped polysilicon
- Normally use transistors in CMOS to implement resistor and current-source loads
- If biased for use as a resistor, called an unsaturated load inverter
- If load transistor operates in saturation as a constant current source, called a saturated load inverter

Pseudo NMOS Inverter





$V_{DD} + V_{dsp} = V_{out}$

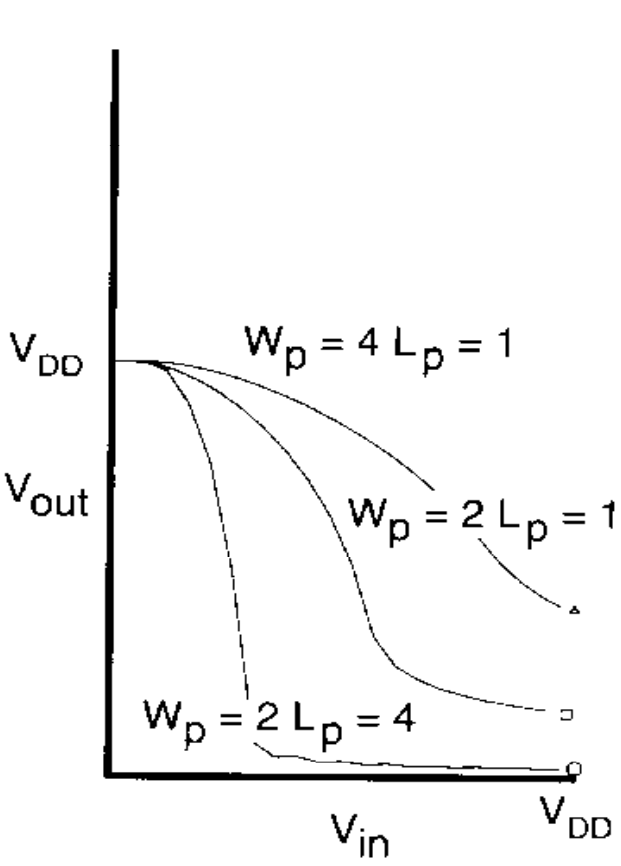
$\Rightarrow V_{dsp} = V_{out} - V_{DD}$

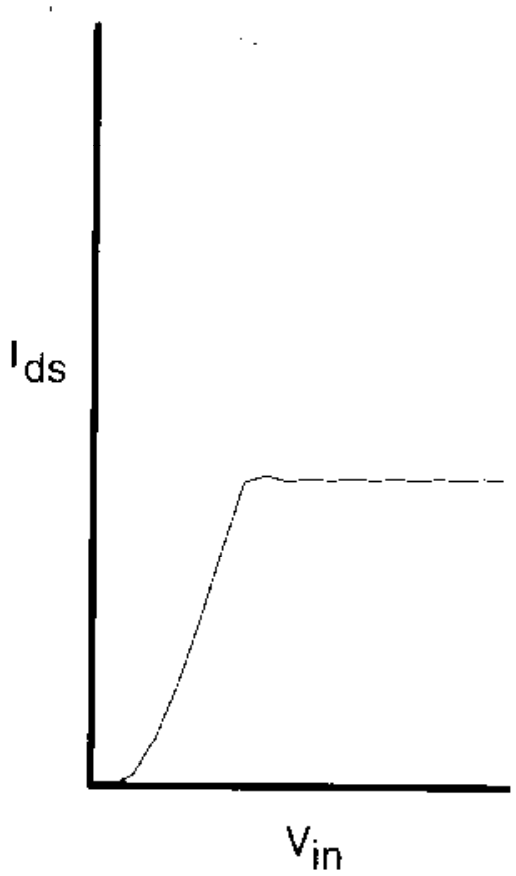
$\Rightarrow V_{dsp} = V_{out} + V_{gsp}$

$\therefore V_{dsp} > V_{gsp} - V_{tp}$ or $V_{out} > - V_{tp}$

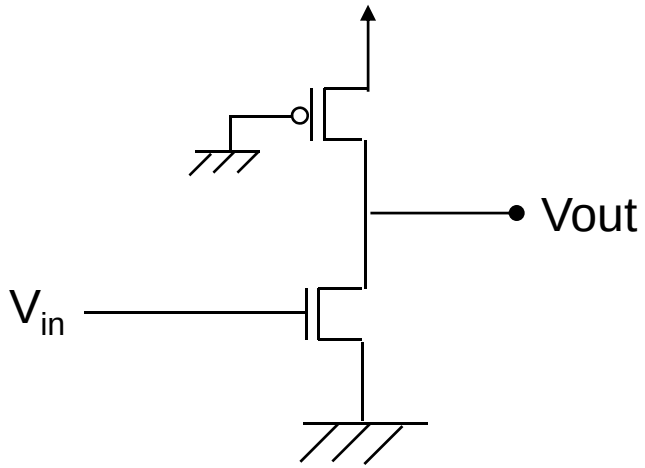
$\Rightarrow$ Non-saturated region

DC Transfer Characteristics



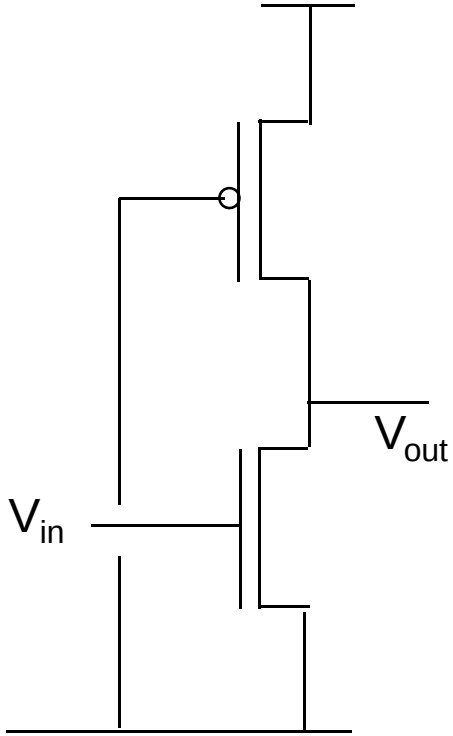


Pseudo-NMOS Inverter



- DC current flows when the inverter is turned on unlike CMOS inverter
- CMOS is great for low power unlike this circuit (e.g. watch needs low power lap-tops etc)
- Need to be turned off during IDDQ (V_{DD} Supply Current Quiescent) testing

PMOST Load with Constant V_{GS}



$V_{oh} = 5.0V$

$V_{ol} = ???$

$I = 0.5\beta_p.(V_{dd}-V_{tp})^2$

$I = \beta_n.((V_{dd}-V_{tn})V_{ol}-0.5V_{ol}^2)$

$$\frac{\beta_n}{\beta_p} = \frac{0.5(V_{dd} - V_{tp})^2}{((V_{dd} - V_{tn})V_{ol} - 0.5V_{ol}^{\ 2})}$$

Sizing for V_{OL}

$$\frac{\beta_n}{\beta_p} = \frac{0.5(V_{dd} - V_{tp})^2}{((V_{dd} - V_{tn})V_{ol} - 0.5V_{ol}^{\ 2})}$$

Assume: $V_{dd} = 5.0V$
$V_{tn} = V_{tp} = 1.0V$

Proper design: $V_{ol} < V_{th}$

Let: $V_{ol} = 0.5V$

$$\frac{\beta_n}{\beta_p} = 4.26$$

Sizing for Gate Threshold Voltage (Trip Point)

N-device: saturated $(V_{out} > V_{in} - V_{tn})$

$$I_{dsn} = \frac{\beta_n}{2}(V_{in} - V_{tn})^2$$

P-device: non-saturated

$$V_{gsp} = -V_{DD}$$

$$I_{dsp} = \beta_p[(-V_{DD} - V_{tp})(V_{out} - V_{DD}) - \frac{(V_{out} - V_{DD})^2}{2}]$$

Equating the two currents we obtain,

$$\frac{\beta_n}{2}(V_{in} - V_{tn})^2 = -\beta_p[(-V_{DD} - V_{tp})(V_{out} - V_{DD}) - \frac{(V_{out} - V_{DD})^2}{2}]$$

Sizing for Gate Threshold Voltage

Solving for V_{out}

$$V_{out} = -V_{tp} + \sqrt{(V_{DD} + V_{tp})^2 - C}$$

Where $C = k (V_{in} - V_{tn})^2$

$$k = \frac{\beta_n}{\beta_p}$$

Also, $$\frac{\beta_n}{\beta_p} = \frac{(V_{DD} + V_{tp})^2 - (V_{out} + V_{tp})^2}{(V_{in} - V_{tn})^2}$$

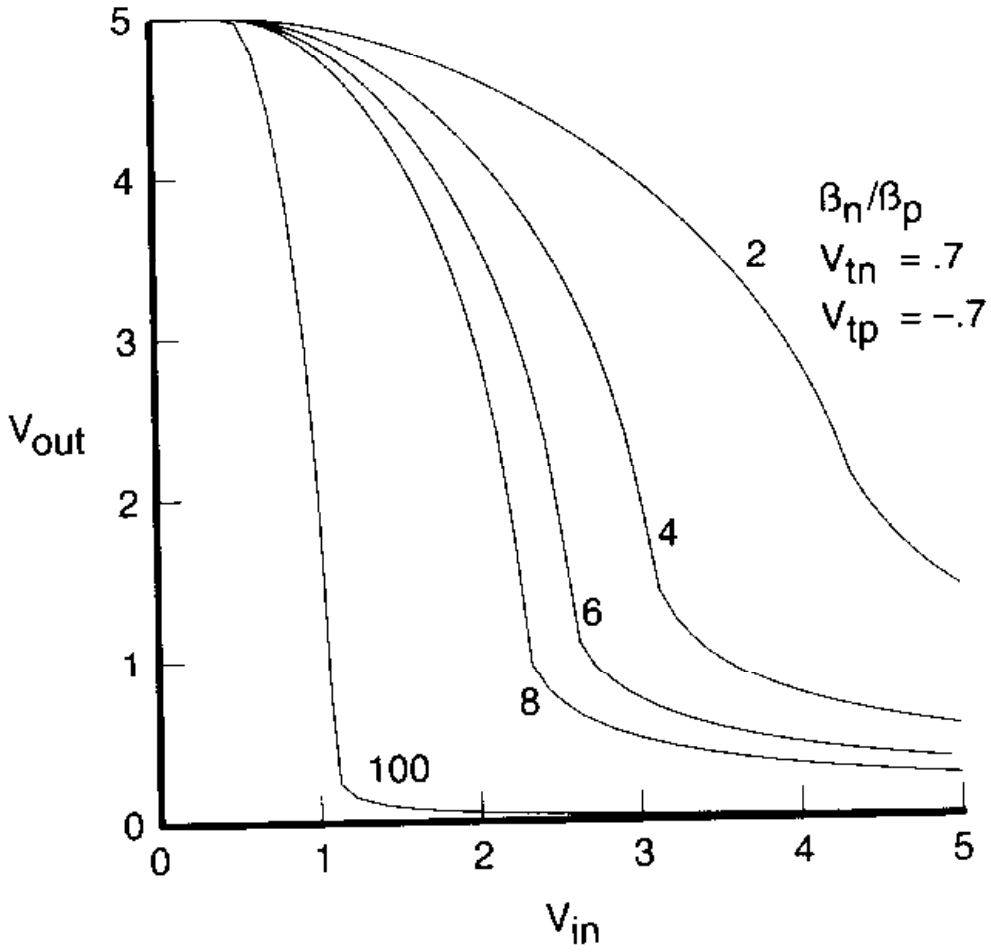
To make gate threshold voltage = $0.5V_{DD}$

$$\frac{\beta_n}{\beta_p} = 6.11$$

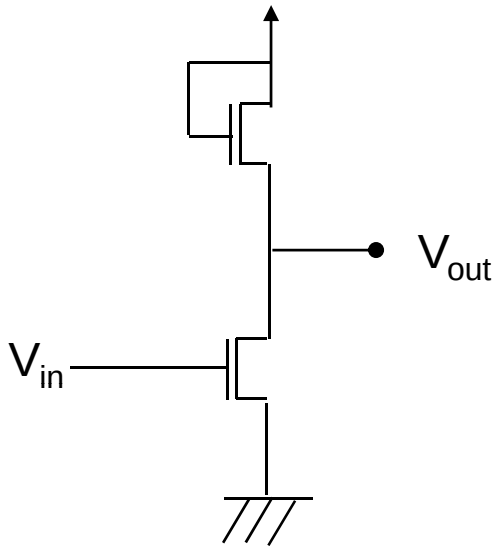
Noise Margin

β_n/β_p	V_{IL}	V_{IH}	V_{OL}	V_{OH}	NM_L	NM_H
2	3.4	4.5	1.4	5	2.0	0.5
4	**1.8**	**3.3**	**0.6**	**5**	**1.2**	**2.7**
6	**1.4**	**2.8**	**0.35**	**5**	**1.05**	**3.2**
8	1.1	2.4	0.24	5	0.86	3.6
100	0.5	1.1	0.00	5	0.5	3.9

VTC of Pseudo-NMOS Inverter

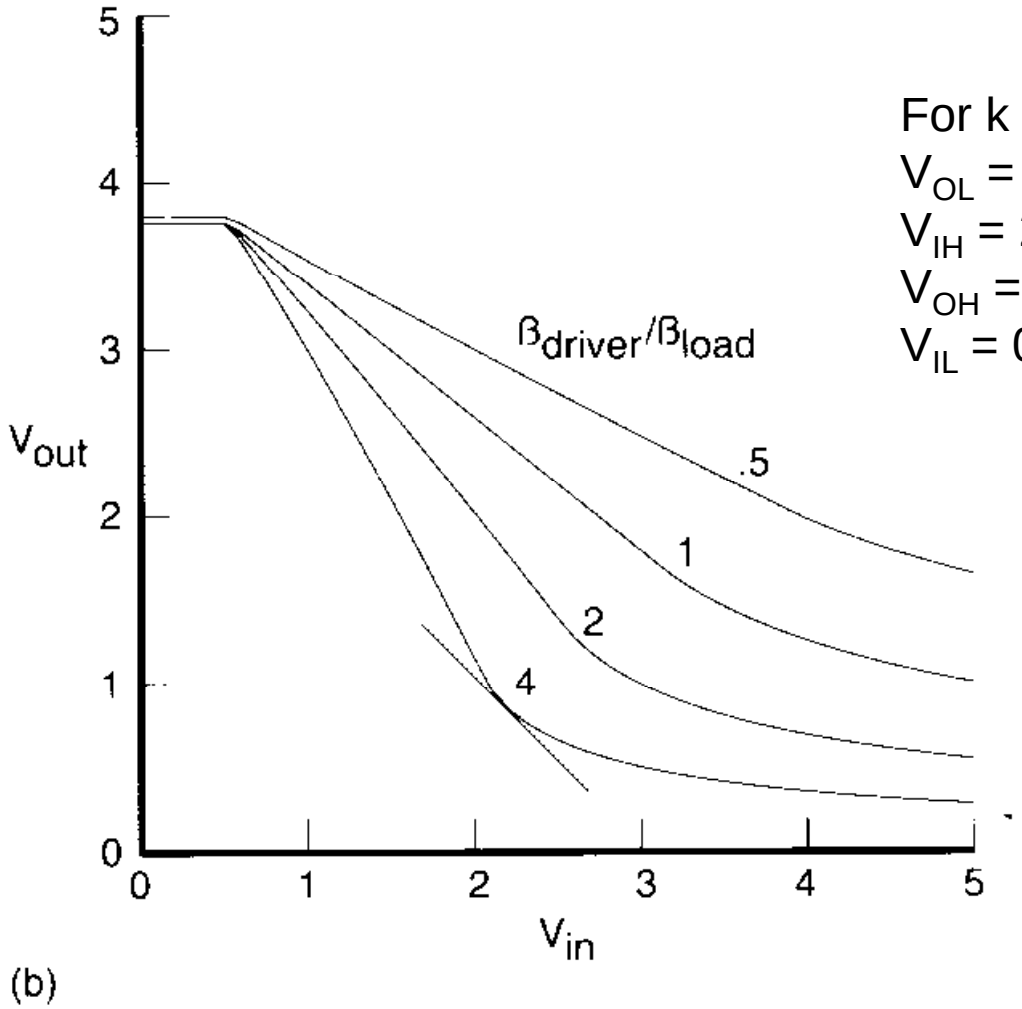


Unsaturated Load Inverter



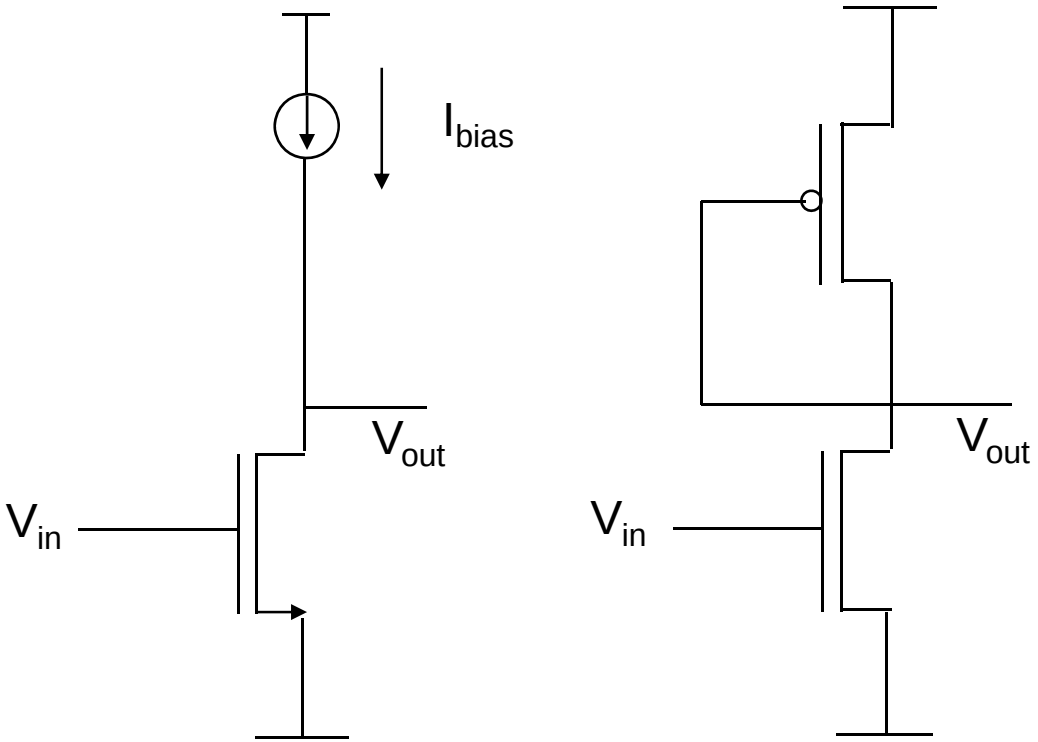
- High is n threshold down from V_{DD}
- Used when depletion mode transistors were not available
- Low noise margin
- Might be used in I/O structures where p-transistors were not wanted

VTC of Unsaturated Load Inverters

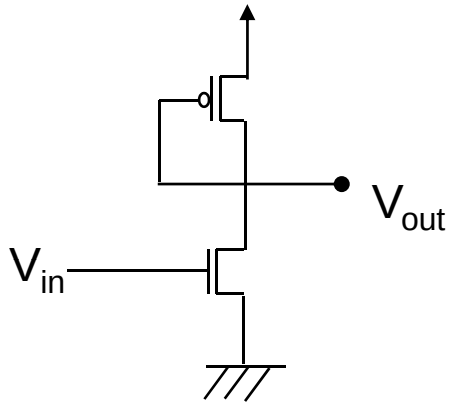


For k = 4
V_{OL} = 0.24V
V_{IH} = 2.2V
V_{OH} = 3.8V
V_{IL} = 0.56V

Current Source Load



Saturated Load Inverter



- $V_{out} > V_{in} - V_{tn} \Rightarrow$ driver transistor in saturation
  - When V_{in} is small
- Load transistor permanently in saturation
  - $V_{dsp} = V_{gsp}$
  - $\therefore V_{dsp} < V_{gsp} - V_{tp}$ or $0 < - V_{tp} \Rightarrow$ Saturated region

When V_{in} is Small

$$I_{ds,driver} = \frac{\beta_{driver}}{2}(V_{in} - V_{tn})^2$$

Load in saturation:

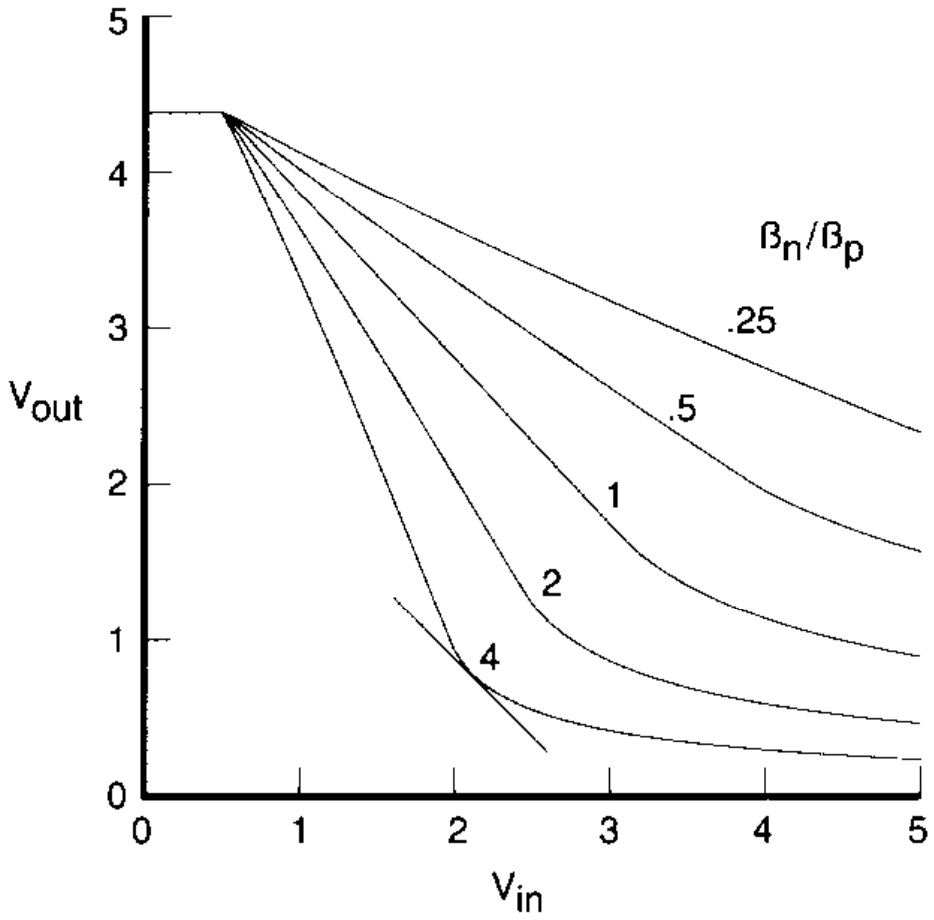
$$I_{ds,load} = -\frac{\beta_{load}}{2}(V_{out} - V_{DD} - V_{tp})^2$$

Equating the currents:

$$V_{out} = V_{DD} + V_{tp} + \sqrt{k}(V_{in} - V_{tn})$$

where $k = \frac{\beta_{driven}}{\beta_{load}}$

VTC of Saturated Load Inverter

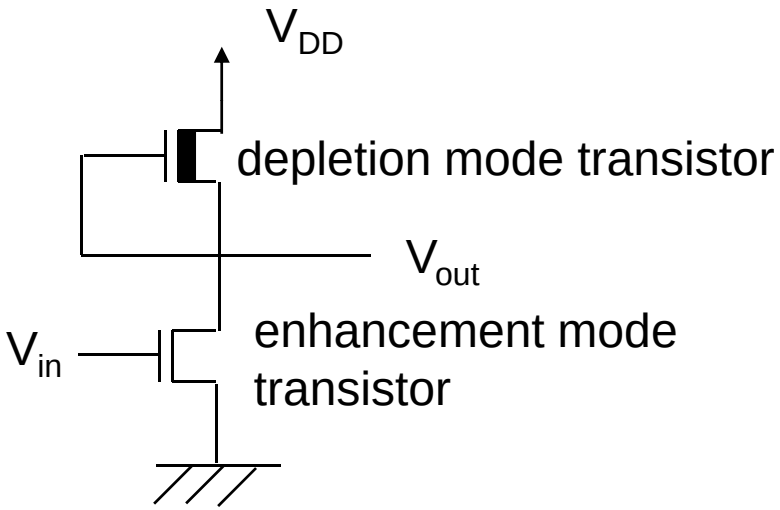


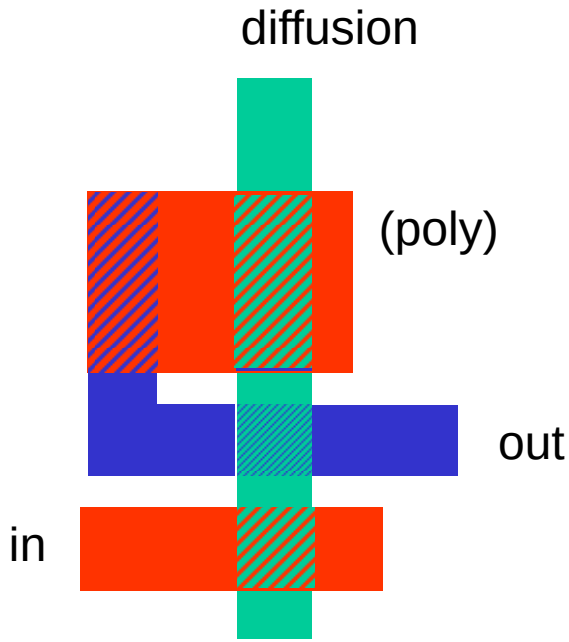
For k = 4
V_{OL} = 0.24V
V_{IH} = 2.1V
V_{OH} = 4.4V
V_{IL} = 0.5V

NMOS Inverter

Use depletion mode transistor as pull-up

V_{tdep} transistor is < 0 V

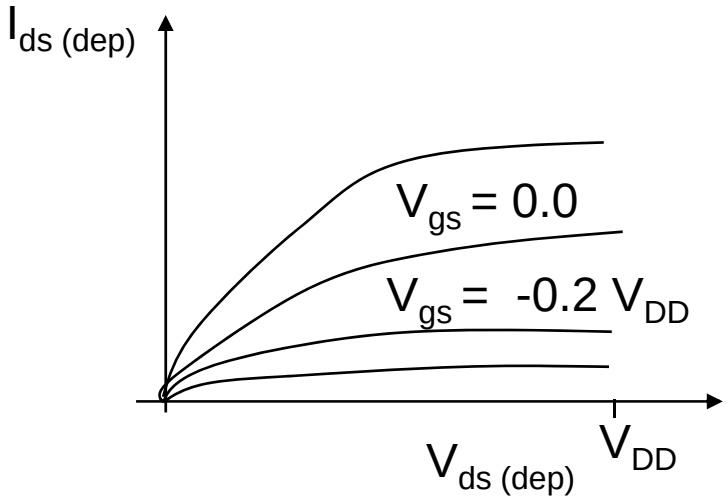


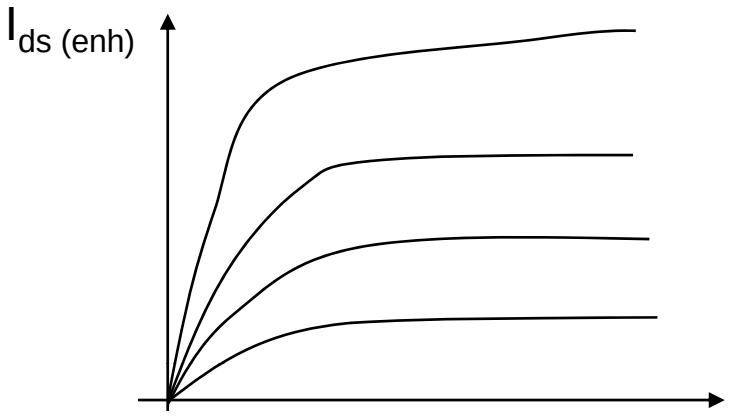


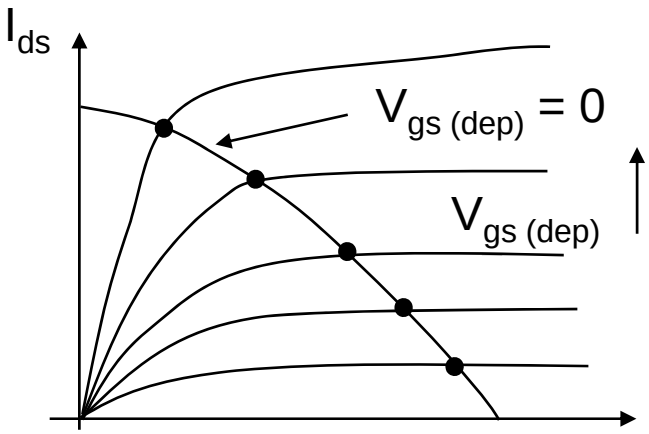
The depletion mode transistor is always ON:
gate and source connected $\Rightarrow V_{gs} = 0$

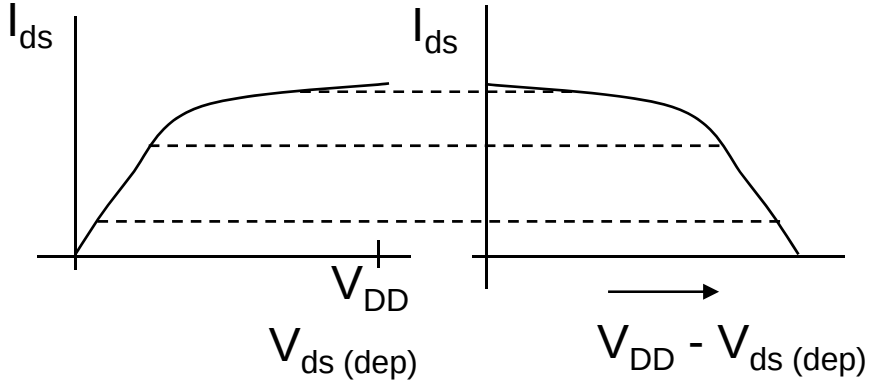
$V_{in} = 0 \Rightarrow$ transistor pull down is off $\Rightarrow V_{out}$ is high

V_{out} vs V_{in} using Graphical Method









$V_{DD} - V_{ds(dep)} = V_{ds(enh)} = V_{out}$

In a steady state,

I_{ds} of both transistors are equal

$V_{ds\ (enh)} = V_{DD} - V_{ds\ (dep)}$

$V_{ds\ (enh)} = V_{out}$

Therefore $V_{out} = V_{DD} - V_{ds\ (dep)}$

Gate Threshold Voltage

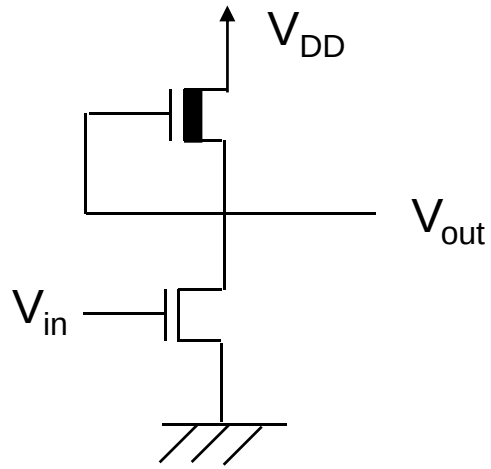
Gate threshold voltage = V_{inv}
= Input voltage at which $V_{in} = V_{out}$

Assume that both driver and load are in saturation with input V_{inv}

$$I_{DS(sat)} = \frac{\beta_{driver}}{2}(V_{gs} - V_t)^2$$

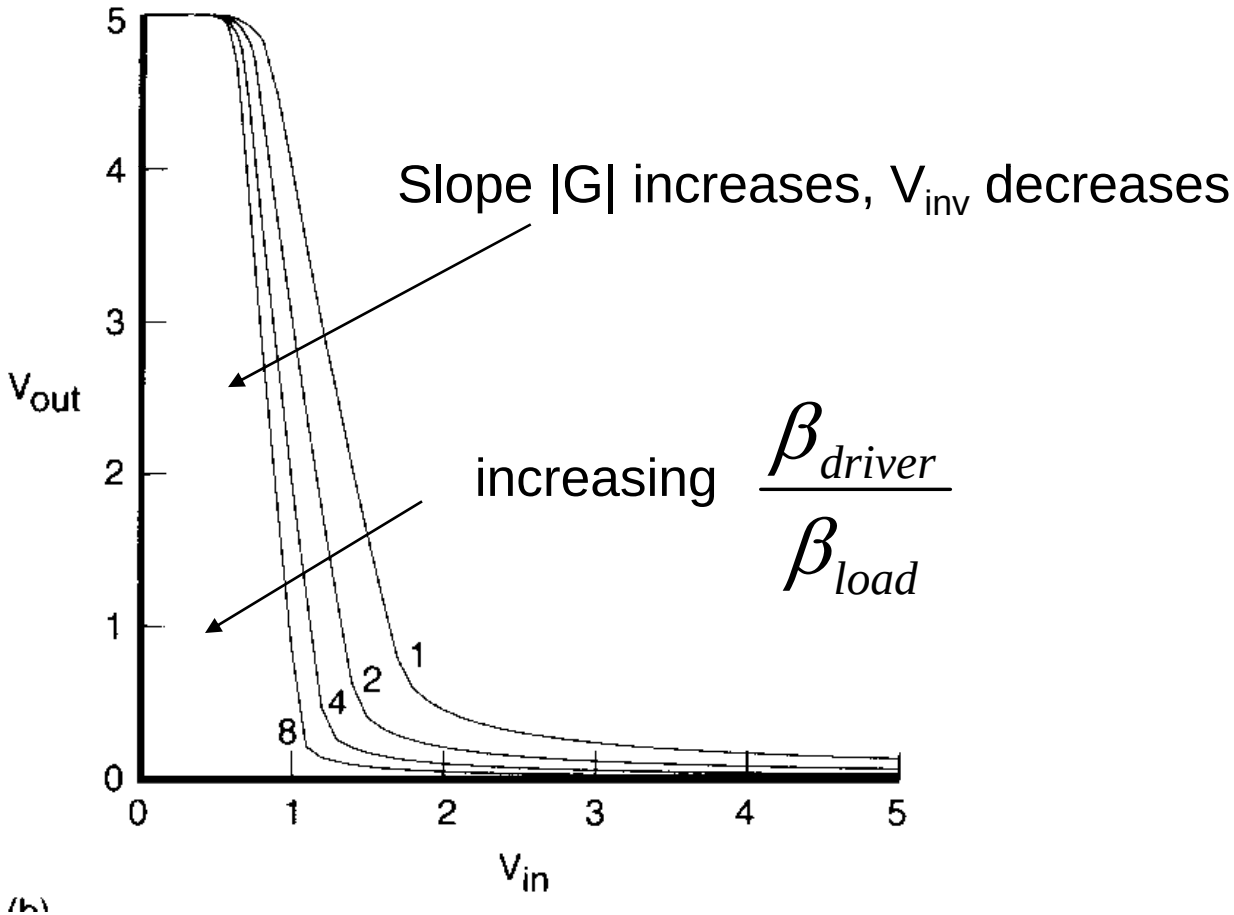
$$\therefore \frac{\beta_{driver}}{2}(V_{inv} - V_t)^2 = \frac{\beta_{load}}{2}(-V_{dep})^2$$

Hence, $$V_{inv} = V_t - V_{dep}\sqrt{\frac{\beta_{load}}{\beta_{driver}}}$$



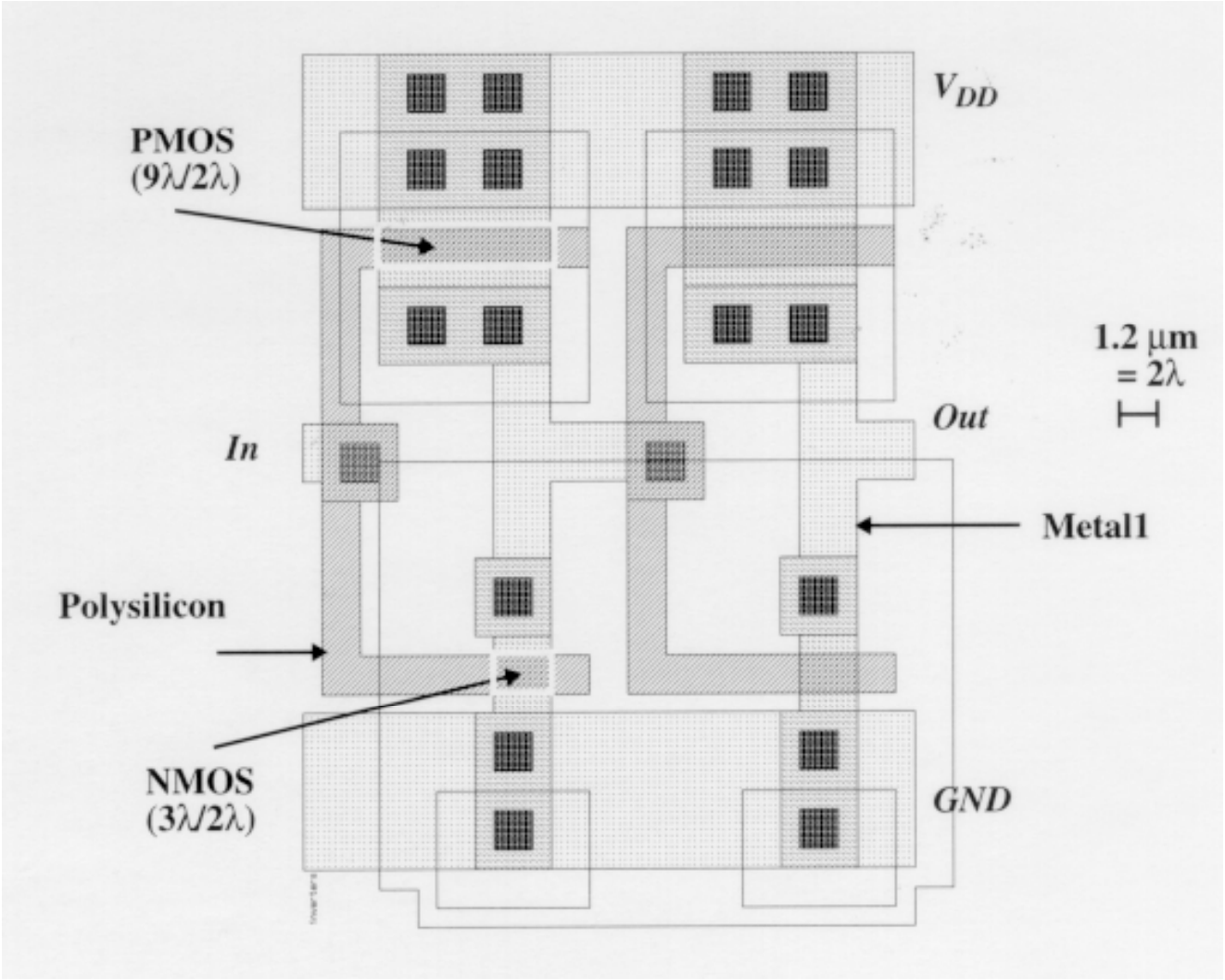
If β_{driver} is increased relative to β_{load} then,
V_{inv} decreases

VTC of NMOS inverter

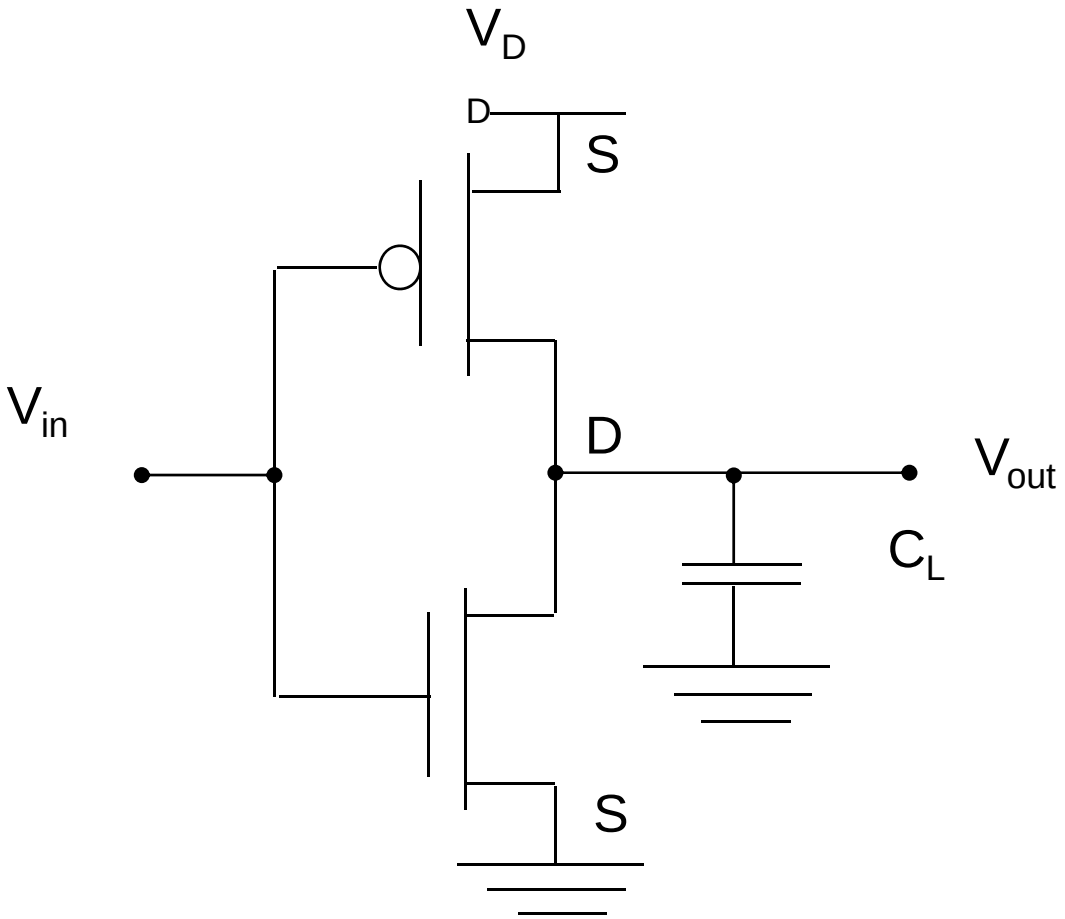


CMOS INVERTER

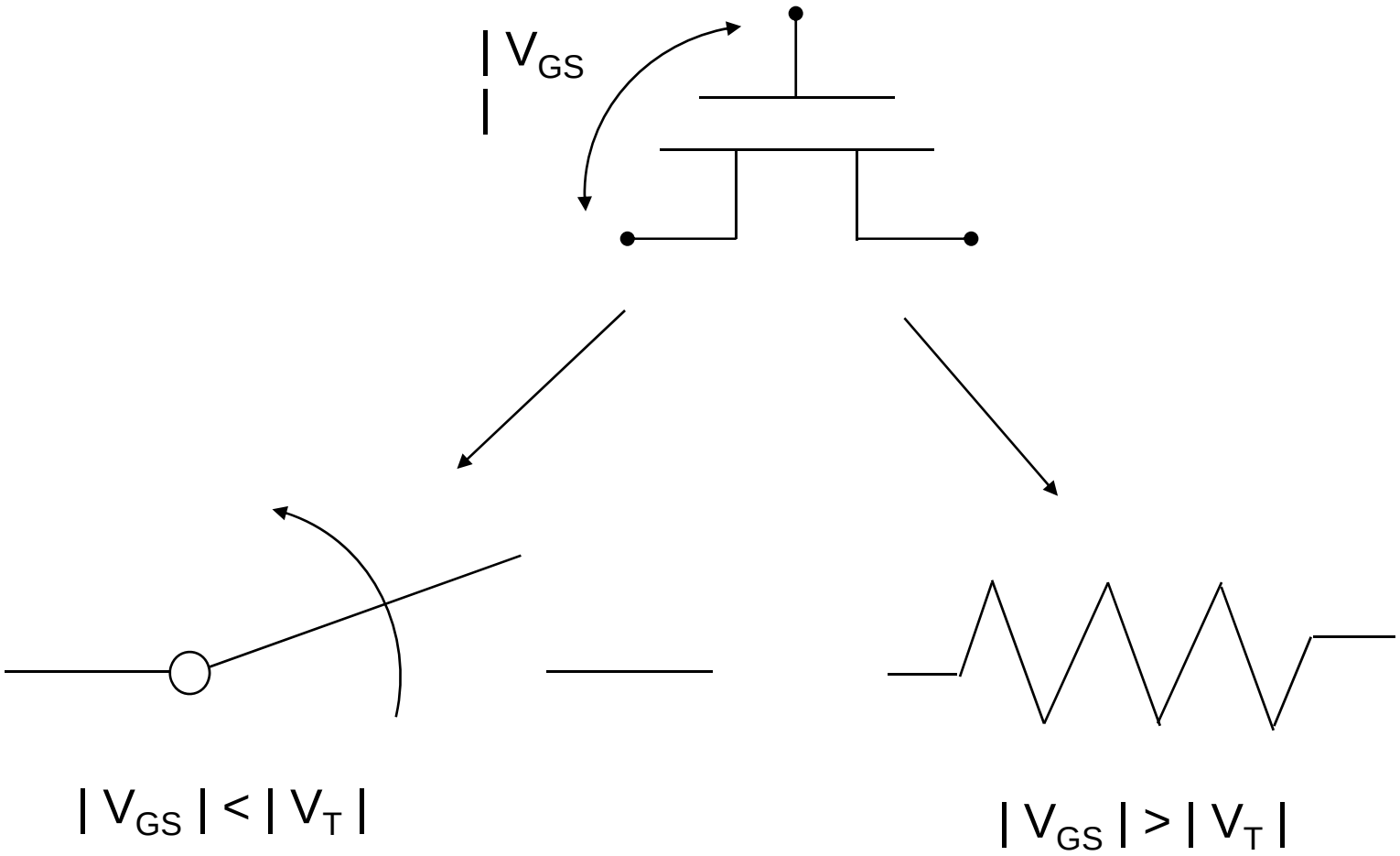
CMOS Inverters



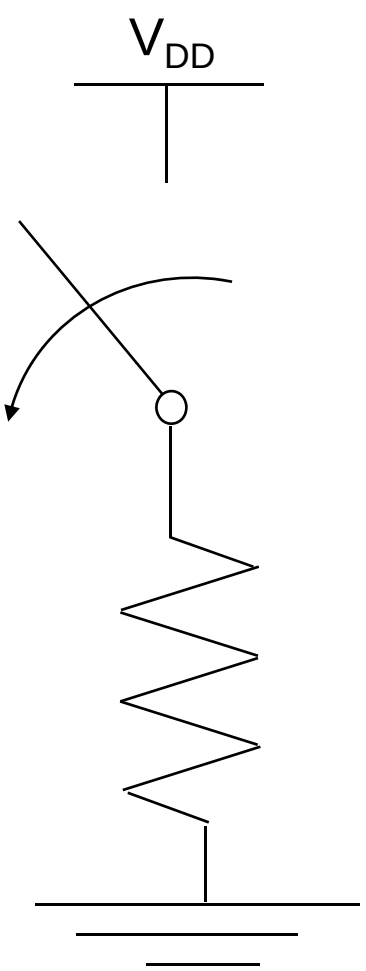
The CMOS Inverter: A First Glance



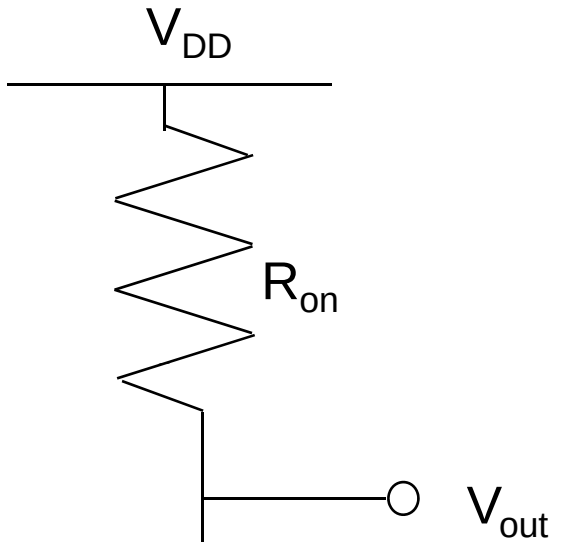
Switch Model of MOS Transistor



CMOS Inverter: Steady State Response

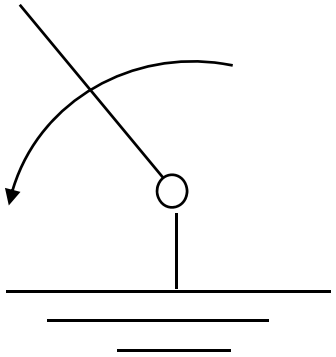


$V_{in} = V_{DD}$



$V_{OH} = V_{DD}$

$V_{OL} = 0$

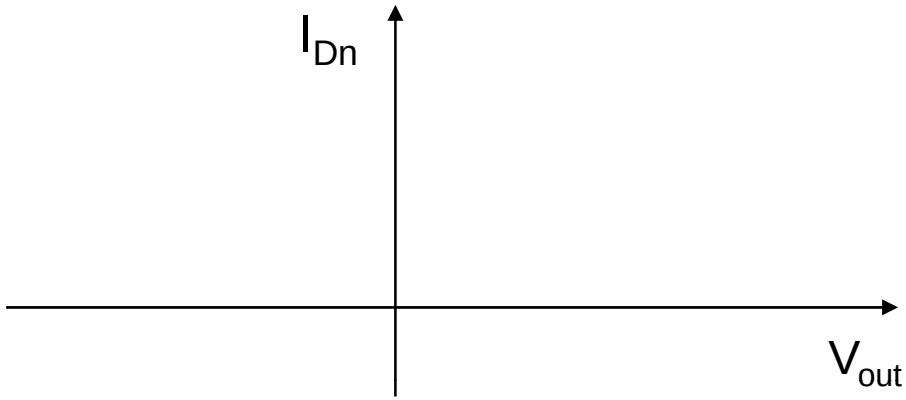


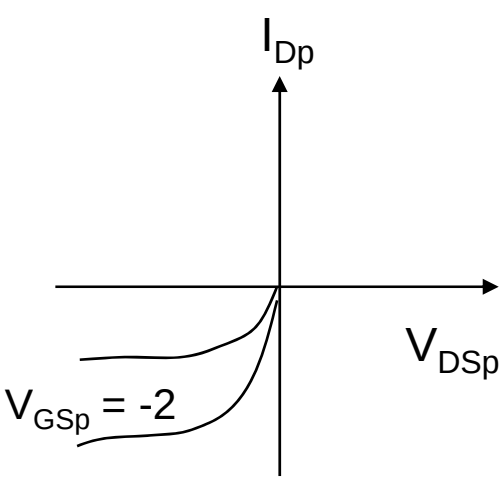
$V_M = f(R_{onn}, R_{onp})$

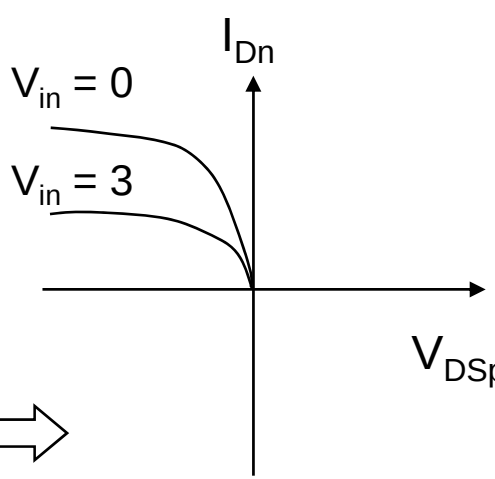
$V_{in} = 0$

PMOS Load Lines

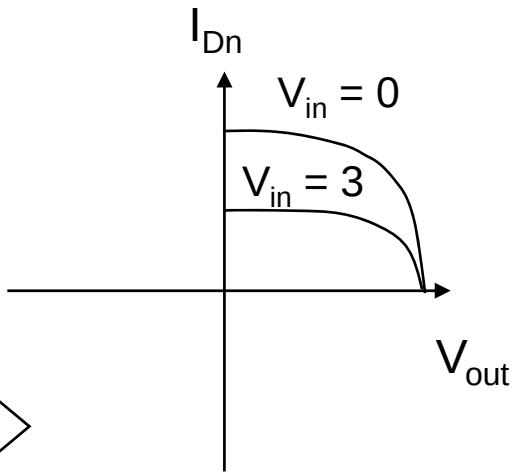
$V_{in} = V_{DD} - V_{GSp}$
$I_{dn} = -I_{DP}$
$V_{out} = V_{DD}-V_{DSp}$





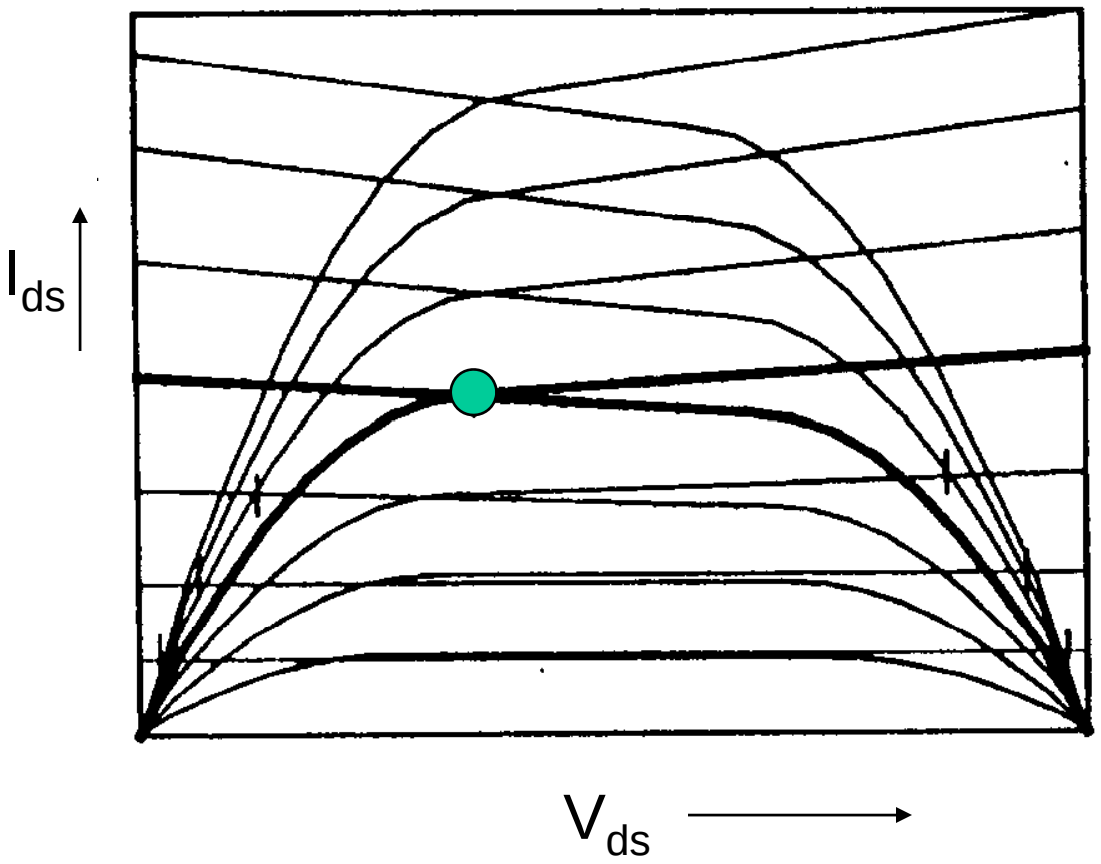


$V_{in} = V_{DD} + V_{GSp}$
$I_{Dn} = - I_{Dp}$

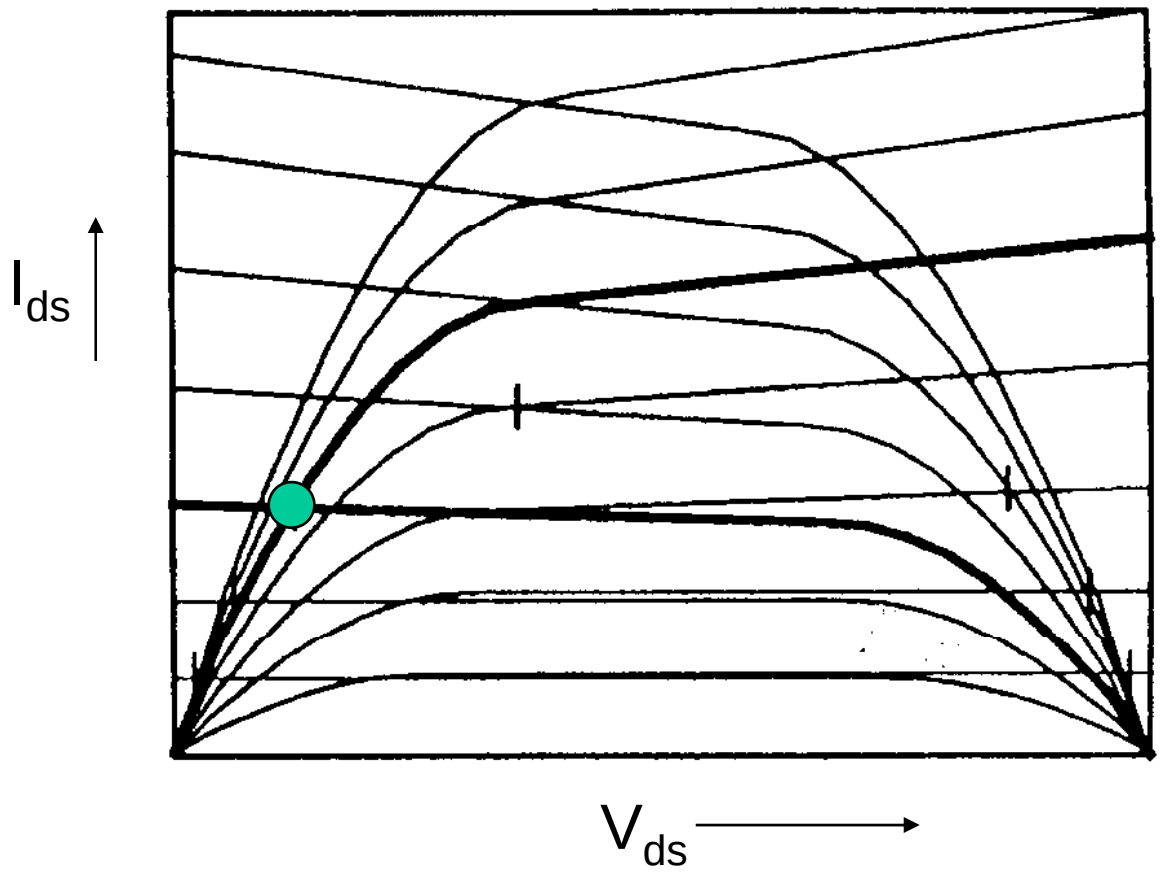


$V_{out} = V_{DD} - V_{DSp}$

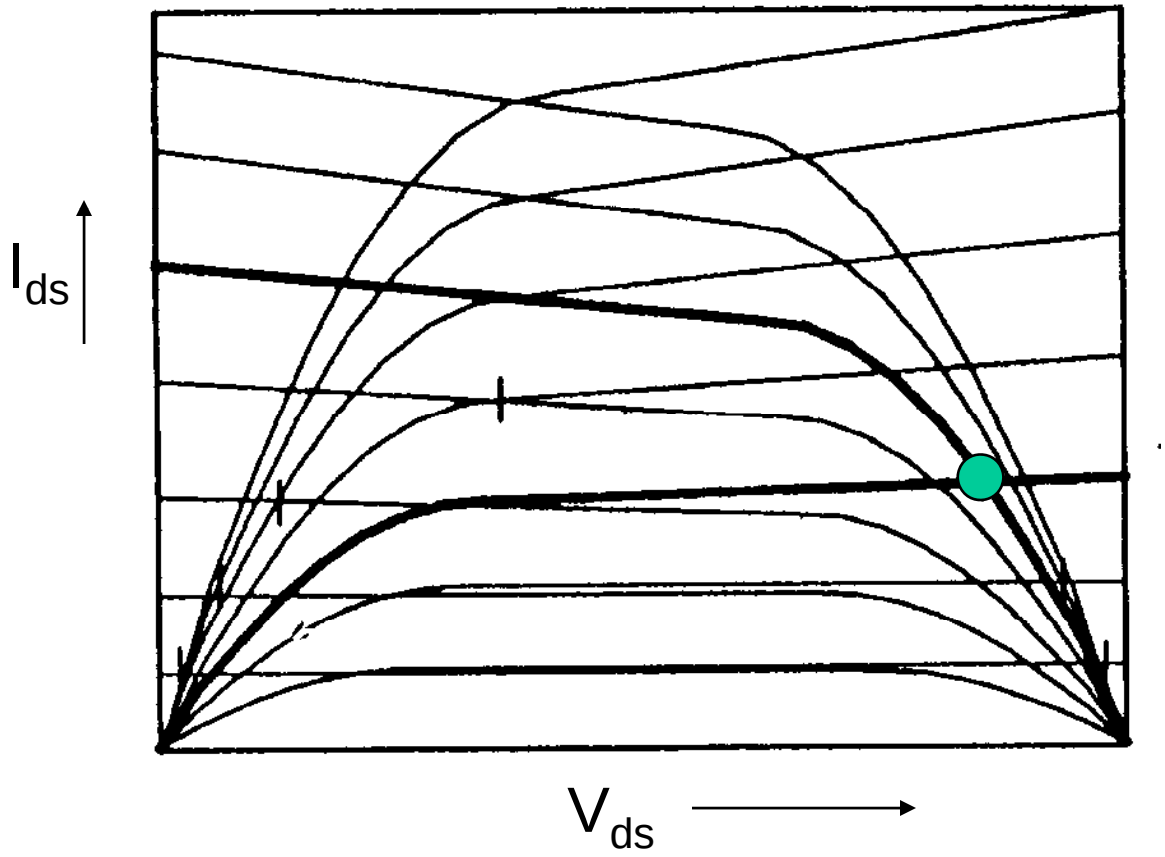
Construction Of Inverter Curves



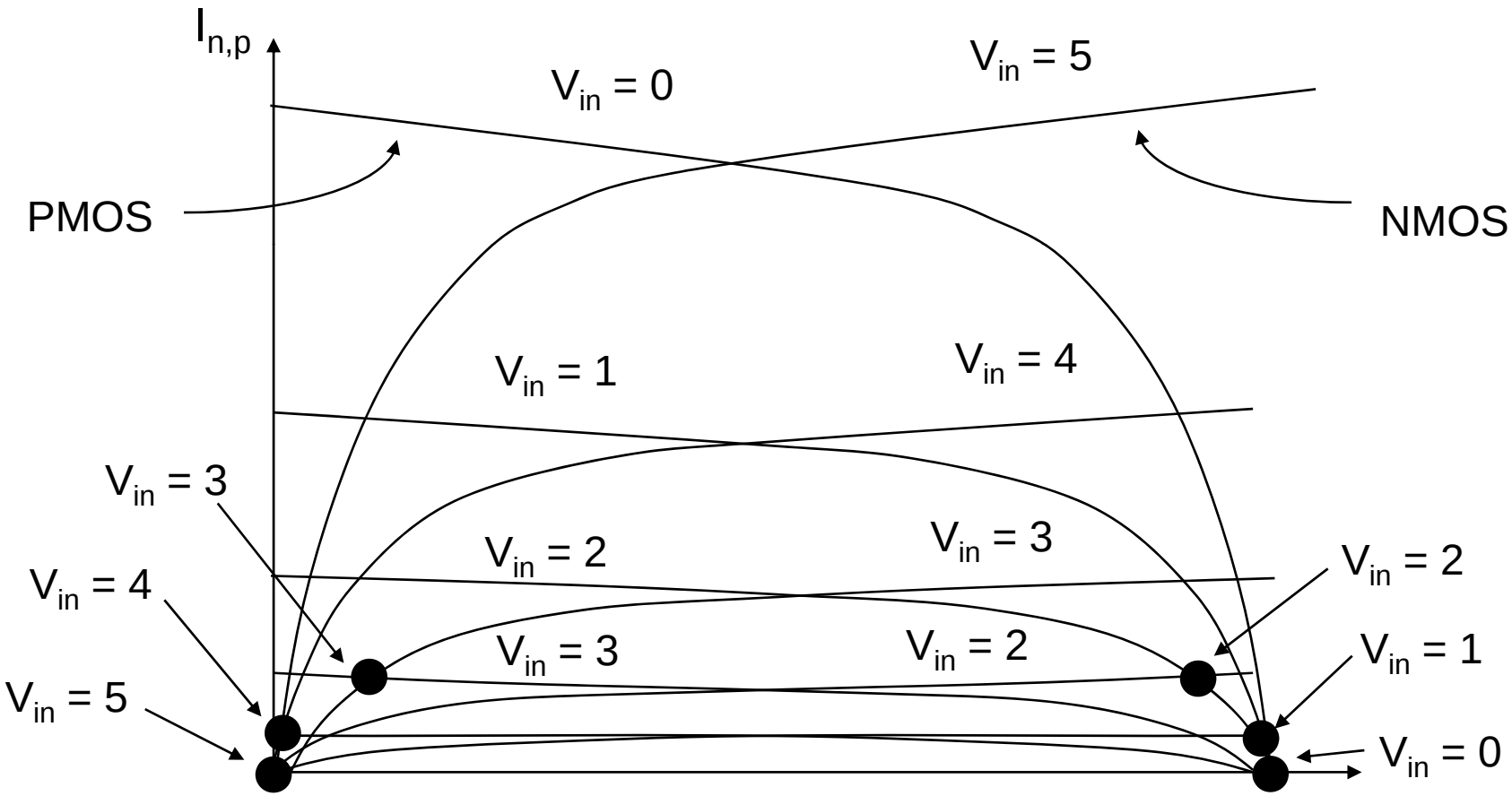
Construction Of Inverter Curves



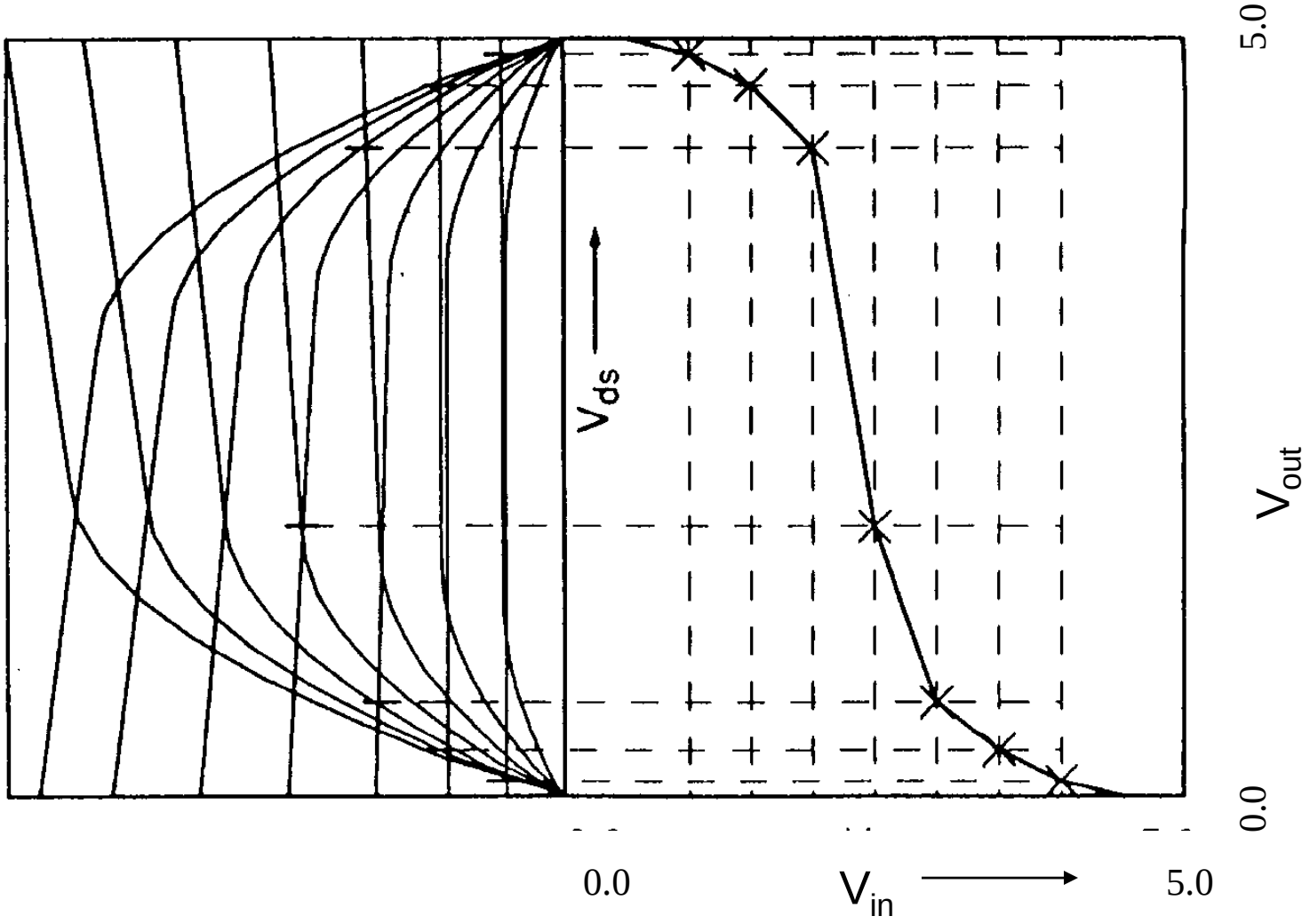
Construction Of Inverter Curves



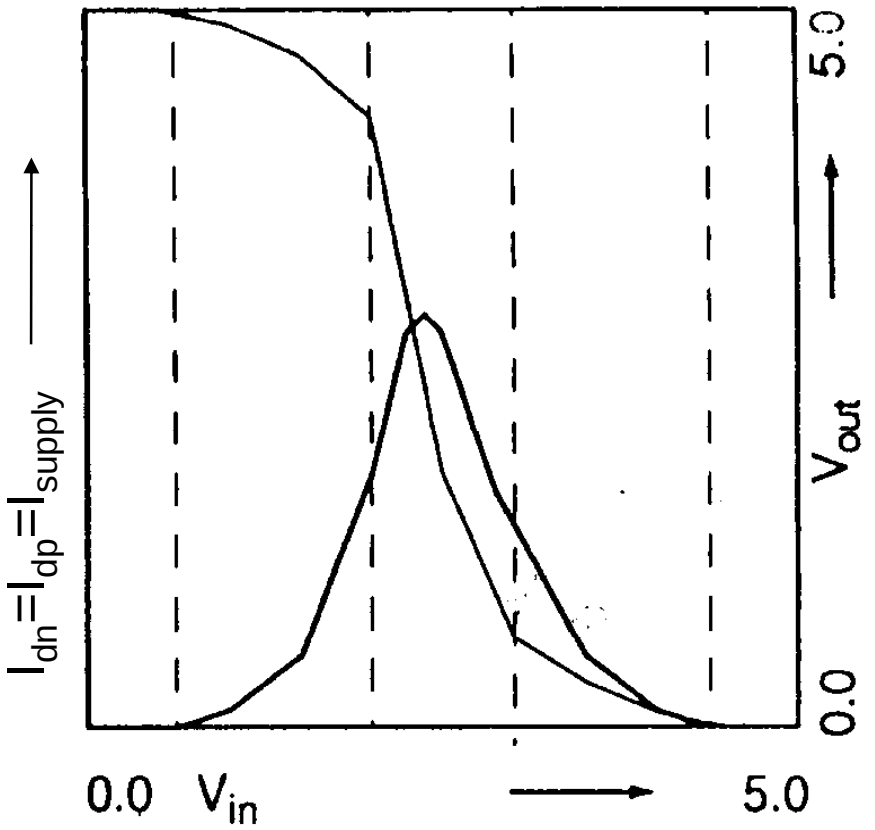
CMOS Inverter Load Characteristics



CMOS Inverter VTC

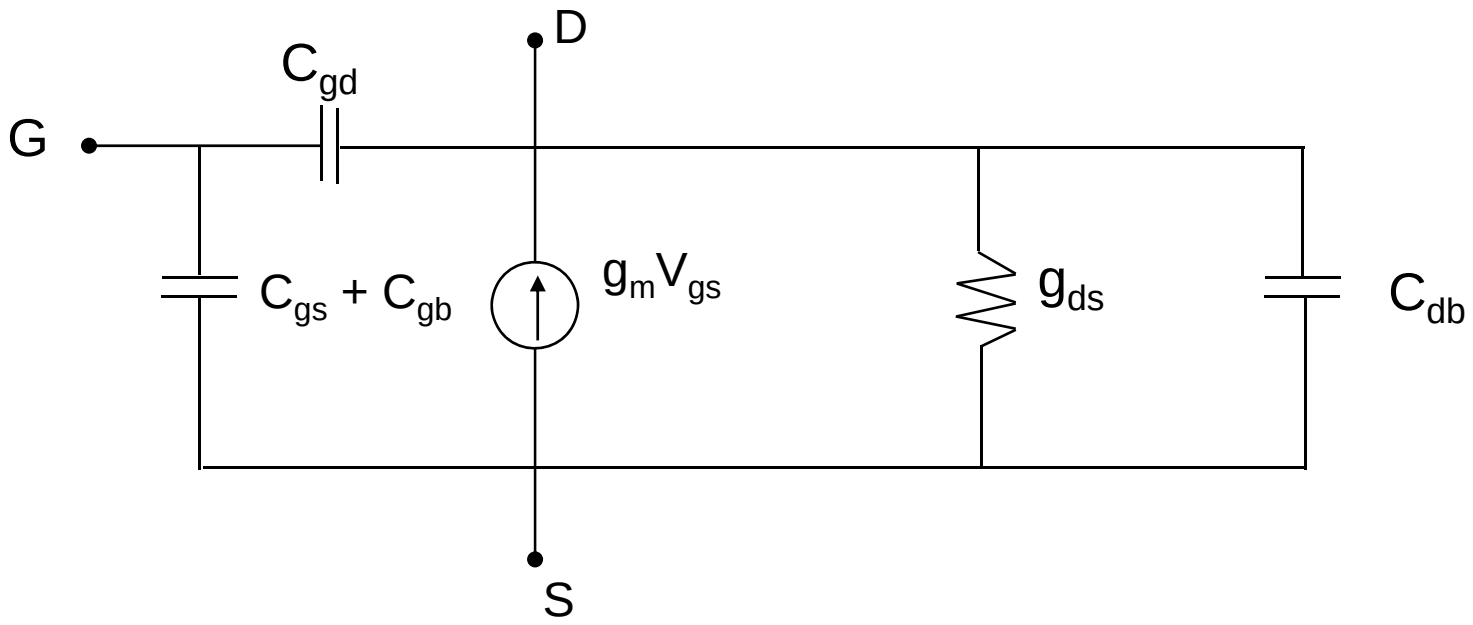


Inverter Supply Current



Small Signal Model for an MOS Transistor

- $V_{sb} = 0$
- voltage-controlled current source (g_m)
- output conductance (g_{ds})
- interelectrode capacitance



Output Conductance

- By differentiating I_{ds} w.r.t. V_{ds}
- In linear region

$$I_{ds} = \beta[(V_{gs} - V_t)V_{ds} - \frac{V_{ds}^2}{2}]$$

$$g_{ds} = \beta[(V_{gs} - V_t) - V_{ds}] \qquad R_{linear} = \frac{1}{\beta(V_{gs} - V_t - V_{ds})}$$

- In saturation, device behaves like a current source: the current being almost independent of V_{ds}

$$I_{ds} = [\frac{\beta}{2}(V_{gs} - V_t)^2]$$

$$\frac{dI_{ds}}{dV_{ds}} = \frac{d[\frac{\beta}{2}(V_{gs} - V_t)^2]}{dV_{ds}} = 0$$

- In reality, secondary effects result in a slope

$$g_{ds} = I_{ds}\lambda$$

Transconductance

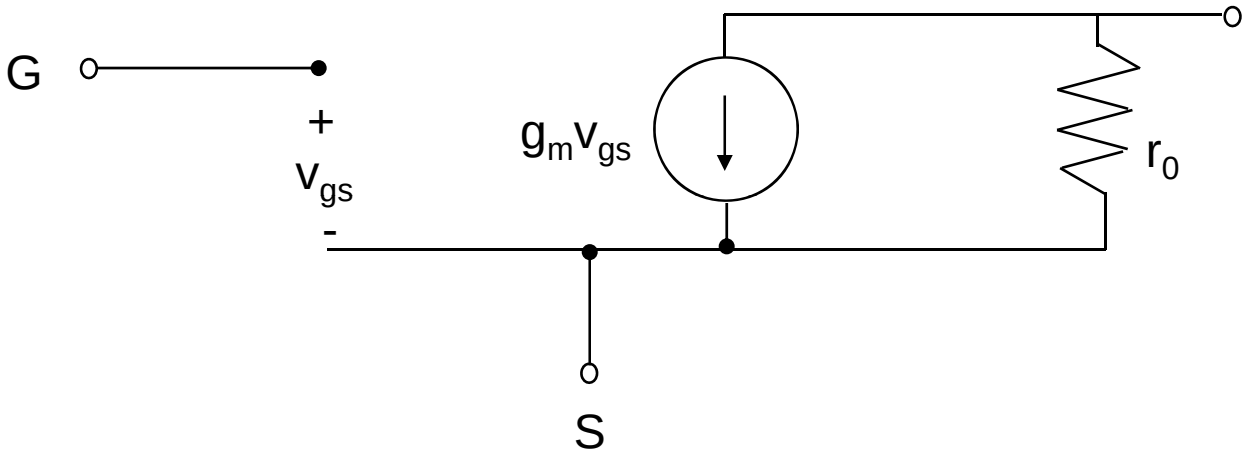
- Expresses relationship between output current and input voltage

$$g_m = \frac{dI_{ds}}{dV_{gs}} \mid V_{ds} = \text{constant}$$

$$g_m(linear) = \beta V_{ds}$$

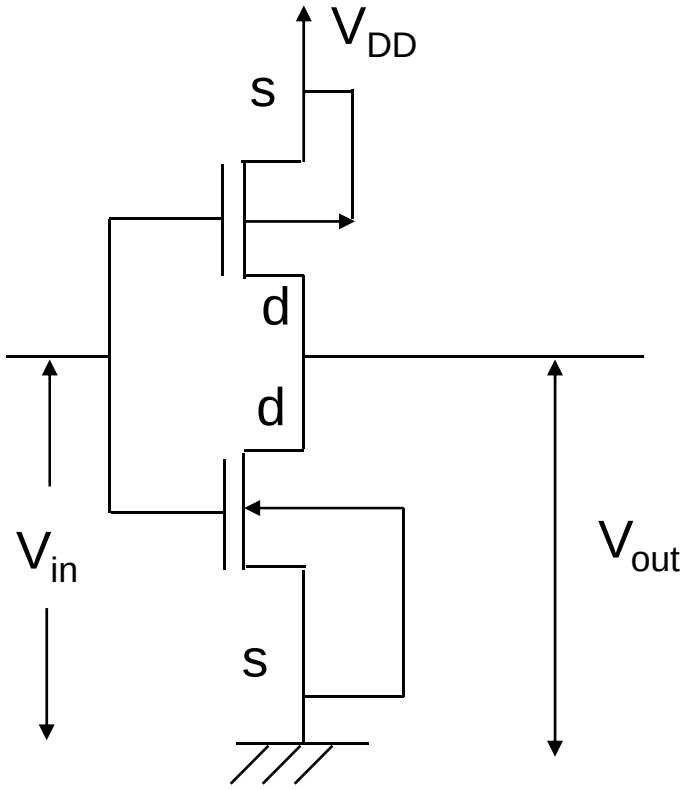
$$g_m(sat.) = \beta(V_{gs} - V_t)$$

MOS Transistor Small Signal Model



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

CMOS Inverter



$$V_{out} = V_{DD} - V_{sdp} = V_{DD} + V_{dsp}$$

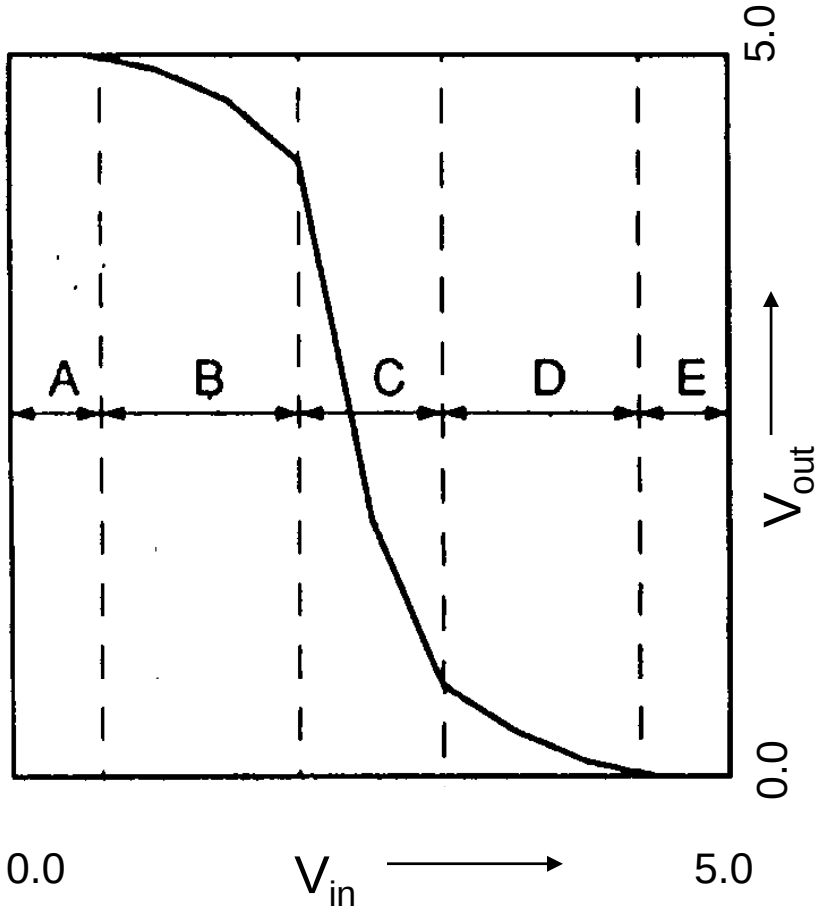
$$V_{in} = V_{DD} - V_{sgp} = V_{DD} + V_{gsp}$$

$$V_{in} = V_{gsn},\ V_{out} = V_{dsn}$$

Regions of Operation

	Cutoff	Non-saturated	Saturated
p-device	$V_{gsp} > V_{tp}$ $V_{in} > V_{tp} + V_{DD}$	$V_{gsp} < V_{tp}$ $V_{in} < V_{tp} + V_{DD}$ $V_{dsp} > V_{gsp} - V_{tp}$ $V_{out} > V_{in} - V_{tp}$	$V_{gsp} = V_{tp}$ $V_{in} < V_{tp} + V_{DD}$ $V_{dsp} < V_{gsp} - V_{tp}$ $V_{out} < V_{in} - V_{tp}$
n-device	$V_{gsn} < V_{tn}$ $V_{in} < V_{tn}$	$V_{gsn} > V_{tn}$ $V_{in} > V_{tn}$ $V_{dsn} < V_{gs} - V_{tn}$ $V_{out} < V_{in} - V_{tn}$	$V_{gsn} > V_{tn}$ $V_{in} > V_{tn}$ $V_{dsn} > V_{gs} - V_{tn}$ $V_{out} > V_{in} - V_{tn}$

Inverter Operating Regions



A: nmost off
pmost linear reg.

B: nmost saturated
pmost linear reg.

C: nmost saturated
pmost saturated

D: nmost linear reg.
pmost saturated

E: nmost linear reg.
pmost off

Inverter Operating Regions

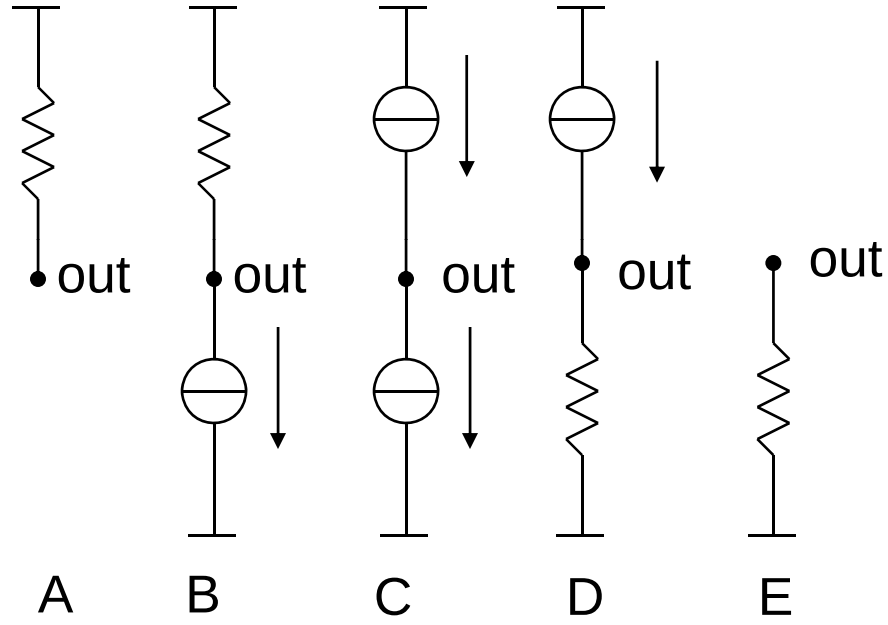
A: nmost off
pmost linear region

B: nmost saturated
pmost linear region

C: nmost saturated
pmost saturated

D: nmost linear region
pmost saturated

E: nmost linear region
pmost off



Assume infinite r_o
when a device is in saturation

Region A

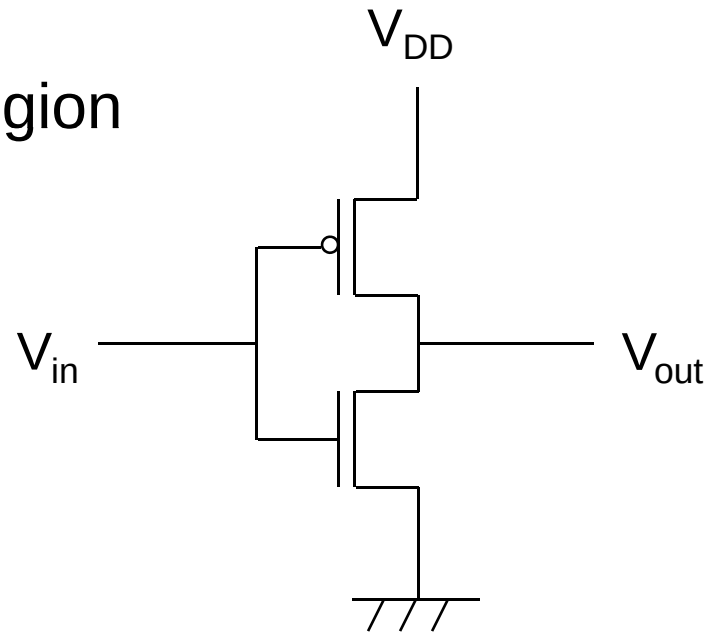
$(0 \leq V_{in} \leq V_{tn})$

$I_{dsn} = 0 \Rightarrow$ n-device is cut-off
p-device in linear region

$I_{dsn} = - I_{dsp} = 0$, as $I_{dsn} = 0$

$V_{dsp} = V_{out} - V_{DD}$

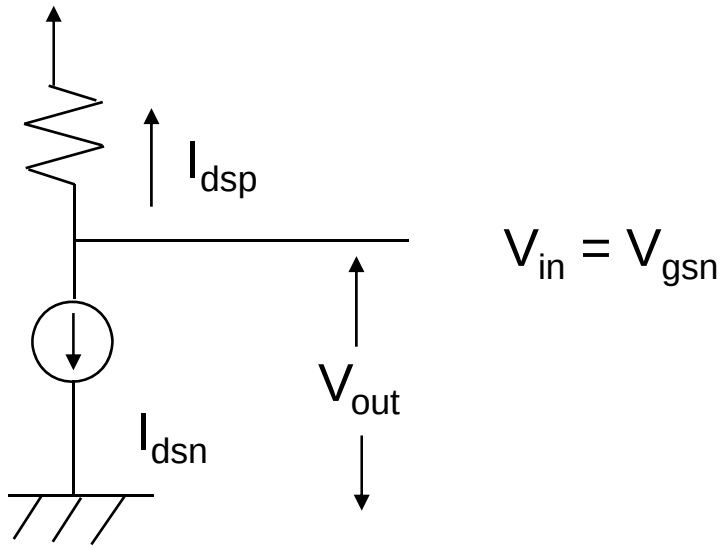
With $V_{dsp} = 0$, $V_{out} = V_{DD}$



Region B

$$(V_{tn} \leq V_{in} \leq \frac{V_{DD}}{2})$$

p-device in non-saturated region ($V_{ds} \neq 0$)
n-device is in saturation



Region B

$$I_{dsn} = \beta_n \frac{[V_{in} - V_{tn}]^2}{2}; \beta = \frac{\mu_n \varepsilon}{t_{ox}} (\frac{W_n}{L_n})$$

$V_{gsp} = (V_{in} - V_{DD})$ & $V_{dsp} = (V_{out} - V_{DD})$

$$\therefore I_{dsp} = -\beta_p [(V_{in} - V_{DD} - V_{tp})(V_{out} - V_{DD}) - \frac{(V_{out} - V_{DD})^2}{2}]$$

$$\beta_p = \frac{\mu_p t}{t_{ox}} (\frac{W_p}{L_p})$$

Equating $I_{dsp} = -I_{dsn}$

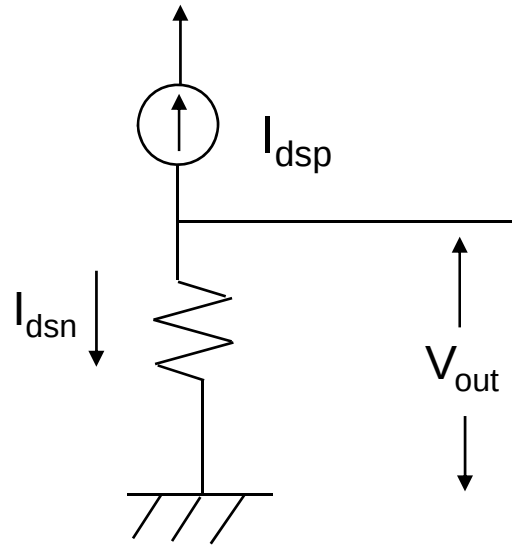
$$V_{out} = (V_{in} - V_{tp}) + \sqrt{(V_{in} - V_{tp})^2 - 2(V_{in} - \frac{V_{DD}}{2} - V_{tp})V_{DD} - \frac{\beta_n}{\beta_p}(V_i n - V_{tc})^2}$$

Region D

$$(\frac{V_{DD}}{2} < V_{in} \le V_{DD} - V_{tp})$$

p : saturation

n : non-saturated



$$I_{dsp} = -\frac{1}{2}\beta_p (V_{in} - V_{DD} - V_{tp})^2$$

$$I_{dsn} = \beta_n [(V_{in} - V_{tn})V_{out} - \frac{V_{out}^{\ 2}}{2}]$$

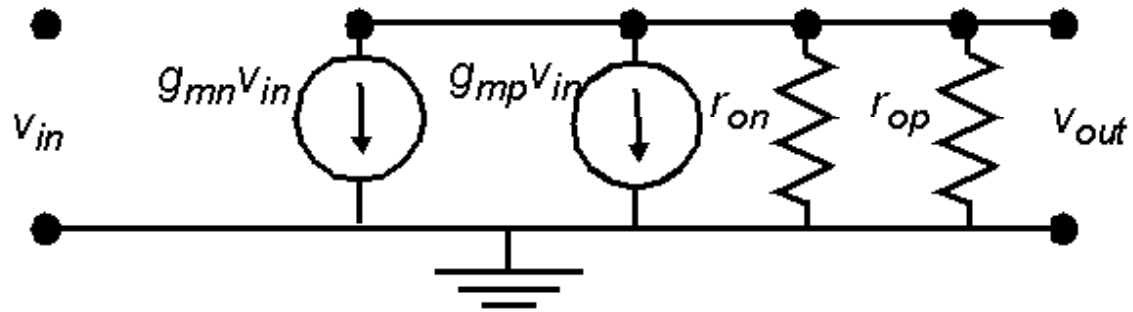
$$I_{dsp} = -I_{dsn}$$

$$\therefore V_{out} = (V_{in} - V_{tn}) - \sqrt{(V_{in} - V_{tn})^2 - \frac{\beta_p}{\beta_n}(V_{in} - V_{DD} - V_{tp})^2}$$

Determining V_{IH} and V_{IL}

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

small-signal model of inverter



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

Region E

(V_{in} >= V_{DD} - V_{tp})

p: cut-off I_{dsp} = 0
n: linear mode

$V_{gsp} = V_{in} - V_{DD} \rightarrow$ more positive than V_{tp}

V_{out} = 0

Region C
(Both devices in Saturation)

$$I_{dsp} = -\frac{\beta_p}{2}(V_{in} - V_{DD} - V_{tp})^2$$

$$I_{dsn} = \frac{\beta_n}{2}(V_{in} - V_{tn})^2$$

Equating $I_{dsp} = -I_{dsn}$

$$V_{in} = \frac{V_{DD} + V_{tp} + V_{tn}\sqrt{\frac{\beta_n}{\beta_p}}}{1 + \sqrt{\frac{\beta_n}{\beta_p}}}$$

Gate Threshold Voltage

If $\beta_n = \beta_p$ & $V_{tn} = -V_{tp}$

$$V_{in} = \frac{V_{DD}}{2}$$

Region C exists for one value of V_{in}

Possible values of V_{out} in region C

n-channel $V_{in} - V_{out} < V_{tn}$

$V_{out} > V_{in} - V_{tn}$

p-channel $V_{in} - V_{out} > V_{tp}$

$V_{out} < V_{in} - V_{tp}$

} saturation conditions

$V_{in} - V_{tn} < V_{out} < V_{in} - V_{tp}$

In reality, region C has a finite slope
- because in reality I_{ds} increases slightly with V_{ds} in saturation

Typical Parameter Values (1μm process)

$$\mu_n = 500 cm^2/V - \sec$$

$$\varepsilon = 3.9 \times 8.85 \times 10^{-14} F/cm$$

$$t_{ox} = 200 \text{Å}$$

$$\beta_n = \frac{\mu\varepsilon}{t_{ox}}\left(\frac{W}{L}\right)$$

$$= \frac{500 \times 3.9 \times 8.85 \times 10^{-14}}{.2 \times 10^{-5}} \frac{W}{L}$$

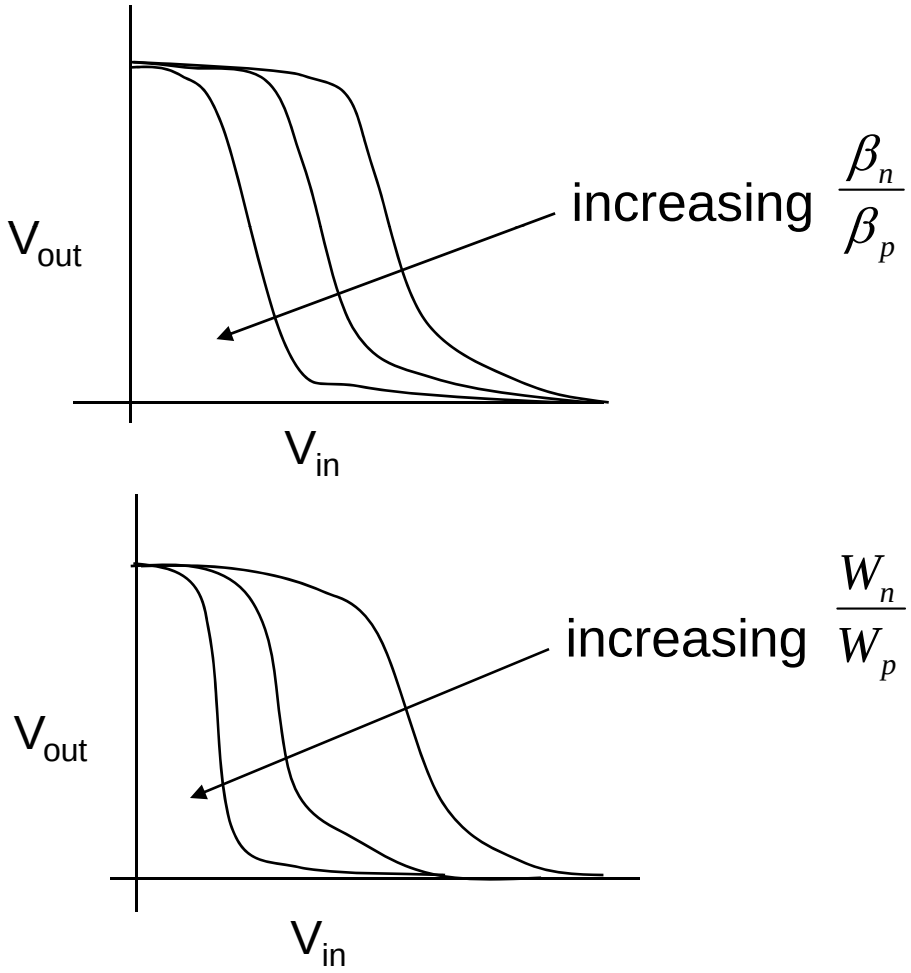
$$= 88.5 \frac{W}{L} \mu A/V^2$$

$$\mu_p \approx 180 cm^2/V - \sec$$

$$\therefore \beta_p = 31.9 \frac{W}{L} \mu A/V^2$$

$$\frac{\beta_n}{\beta_p} = 2.8$$ (The ratio varies from 2-3)

β_n/β_p Ratio



Effect of β_n/β_p Ratio

V_m dependent on $\frac{\beta_n}{\beta_p}$

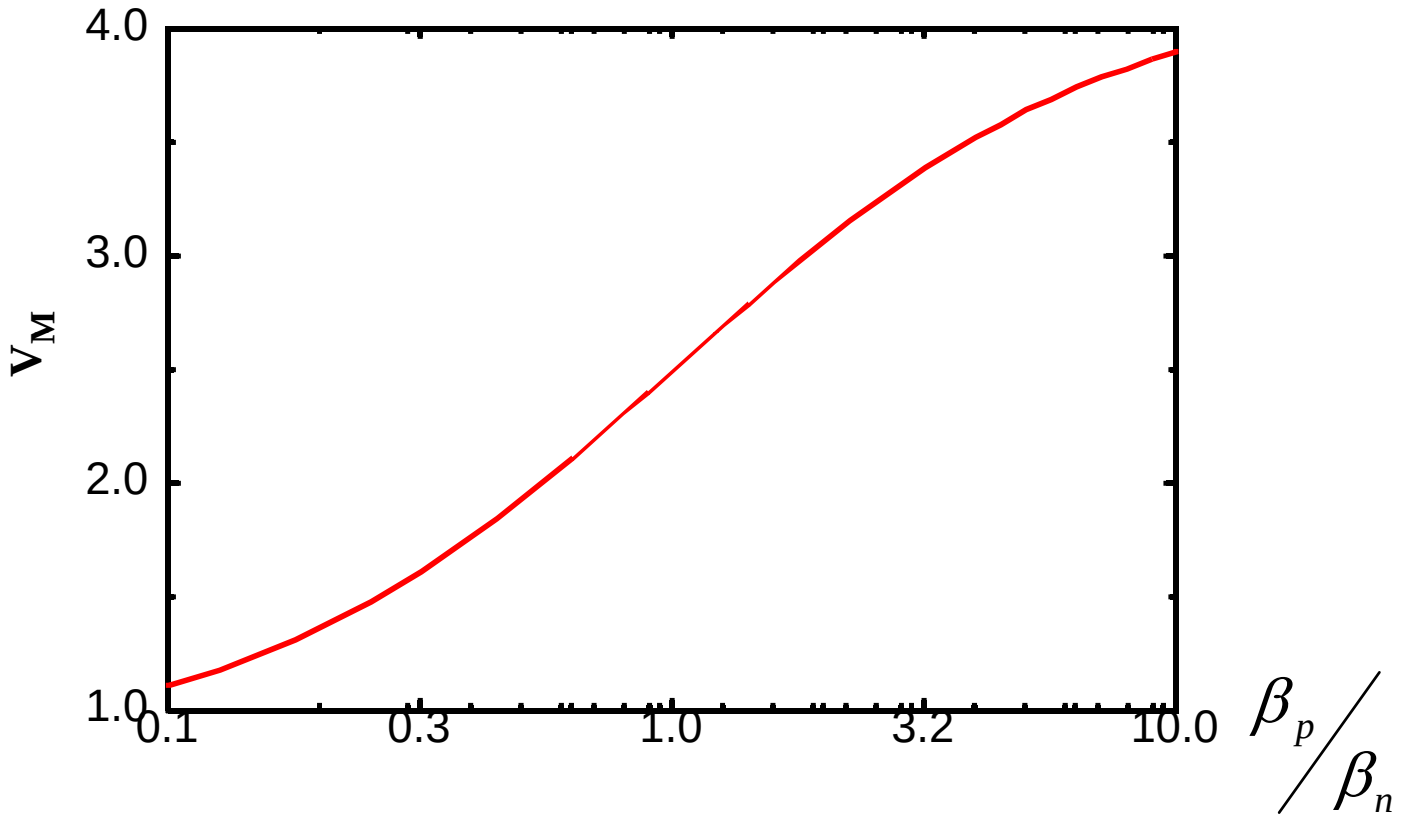
with change in $\frac{\beta_n}{\beta_p}$ transition still remains sharp and hence switching performance does not deteriorate

It is desirable to have

$$\frac{\beta_n}{\beta_p} = 1$$

$\Rightarrow$ allows capacitance load to change and discharge in equal times by providing equal current source & sink capability

Gate Switching Threshold



$$V_M = \frac{r(V_{DD} + V_{tp}) + V_{tn}}{r + 1} \quad \text{with} \quad r = \sqrt{\frac{\beta_p}{\beta_n}}$$

Effect of Temperature

- Temperature similarly affects mobility of holes and electrons
- Temperature increases $\Rightarrow \mu$ decreases $\Rightarrow \beta$ decreases

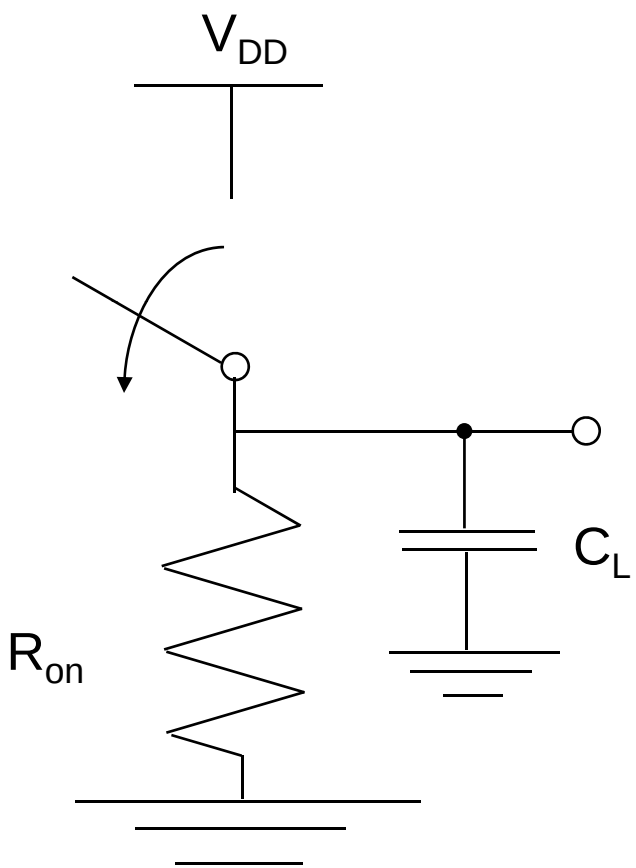
$$\beta \propto T^{-1.5}$$

- Ratio β_n/β_p is independent of temperature to a good approximation
- Temperature, however, reduces threshold voltages
- Extent of region A reduces and extent of region E increases
- VTC shifts to the left as the temperature increases

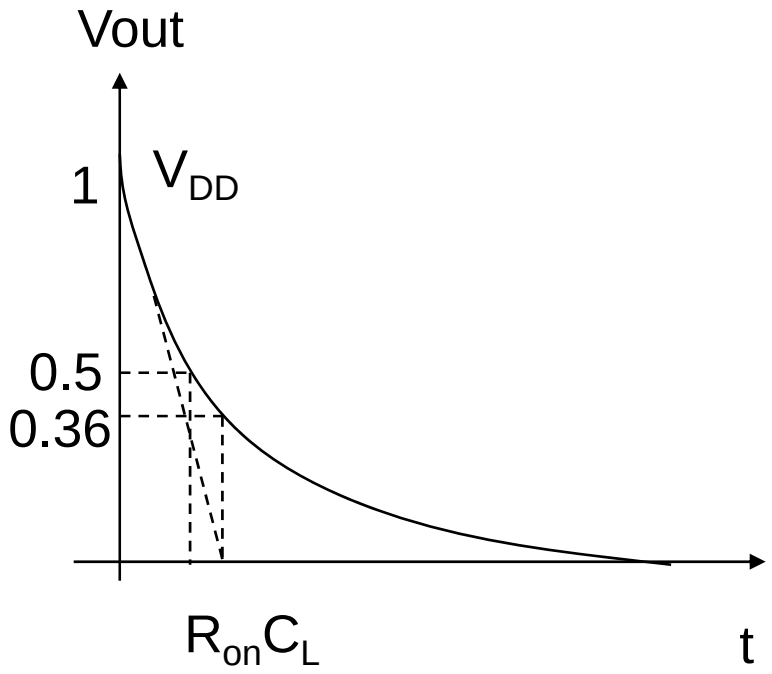
Switching Characteristics

- Switching speed - limited by time taken to charge and discharge, C_L
- Rise time, t_r : waveform to rise from 10% to 90% of its steady state value
- Fall time, t_f : 90% to 10% of steady state value
- Delay time, t_d : time difference between input transition (50%) and 50% output level

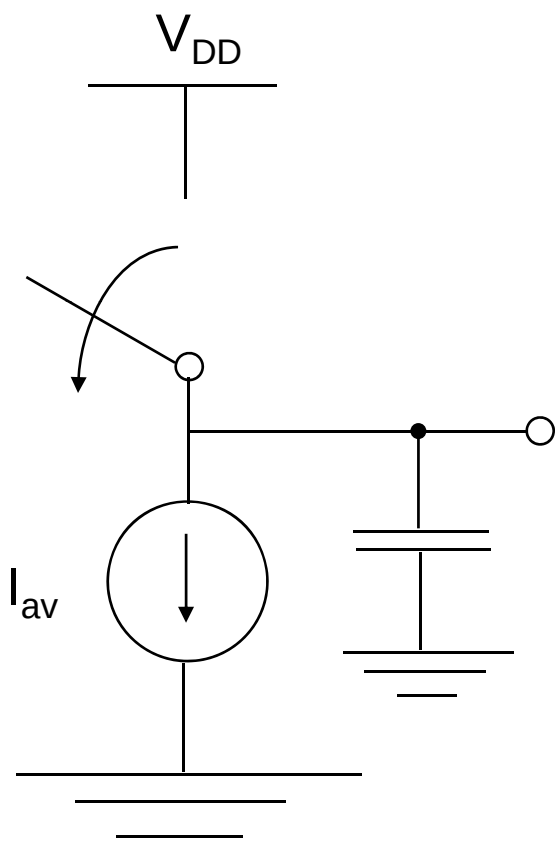
CMOS Inverter: Transient Response



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



CMOS Inverter Propagation Delay



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

$$I_{av} = \frac{I(V_{out} = V_{DD}) + I(V_{out} = V_{DD}/2)}{2}$$

$$= \frac{\beta_n}{2}\left(\frac{7V_{DD}^2}{8} + \frac{V_{tn}^2}{2} - \frac{3V_{DD}V_{tn}}{2}\right)$$

$V_{in} = V_{DD}$

Inverter Propagation Delay

- Assume n-device still in saturation at $V_{out} = V_{DD}/2$

$$I_{av} = \frac{\beta_n}{2}(V_{DD} - V_{tn})^2$$

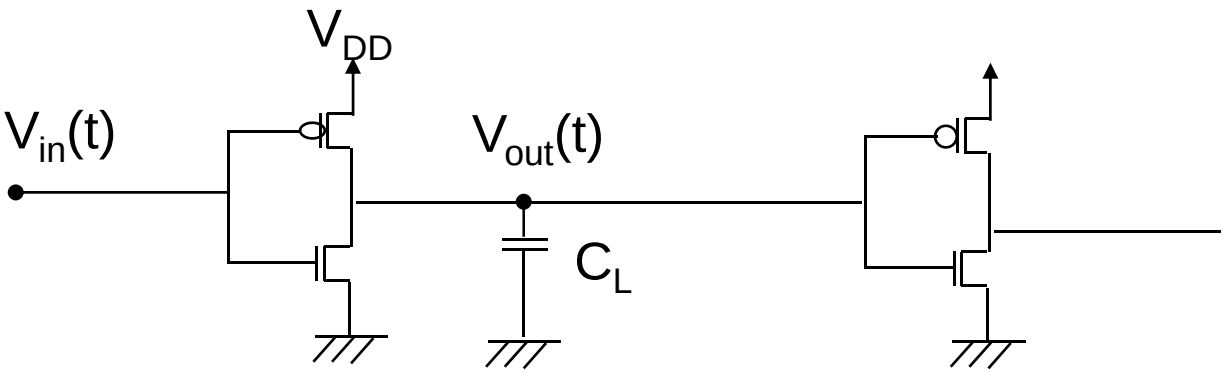
$$t_{pHL} = \frac{C_L V_{DD}}{\beta_n (V_{DD} - V_{tn})^2}$$

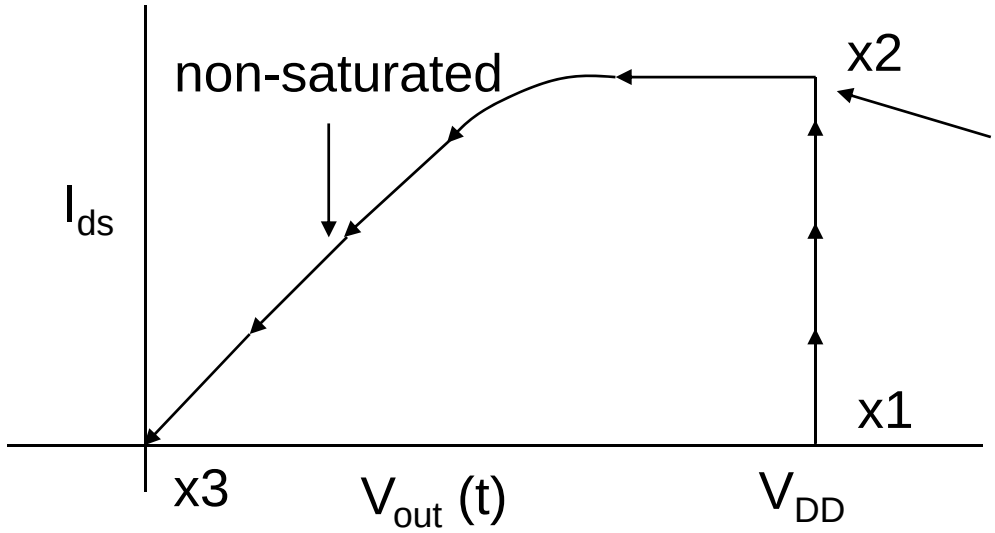
$$\approx \frac{C_L}{\beta_n V_{DD}}$$

$$t_{pLH} \approx \frac{C_L}{\beta_p V_{DD}}$$

$$t_p \approx \frac{C_L}{2V_{DD}}\left(\frac{1}{\beta_p} + \frac{1}{\beta_n}\right)$$

Analysis of Fall Time





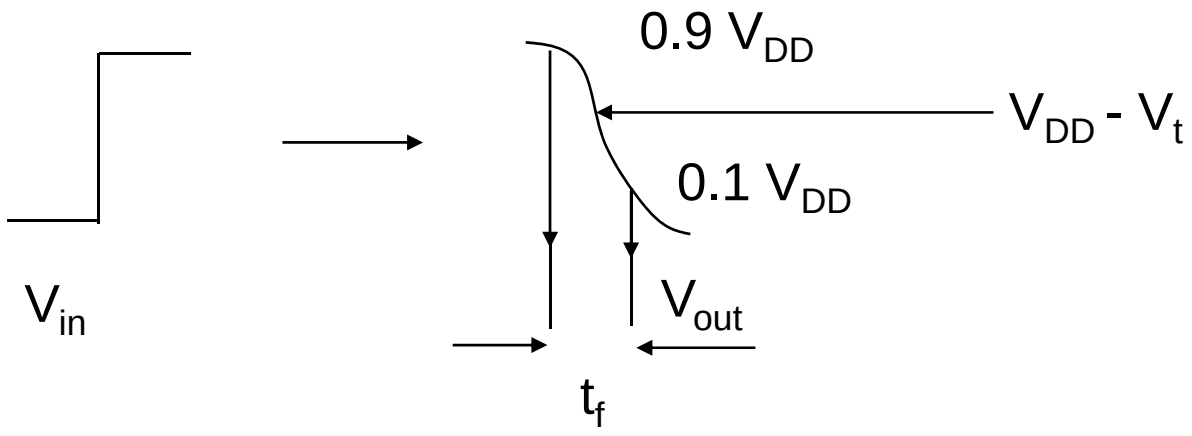
saturated
($V_{ds} = V_{gs} - V_t$)

Application of step input

Components of Fall Time

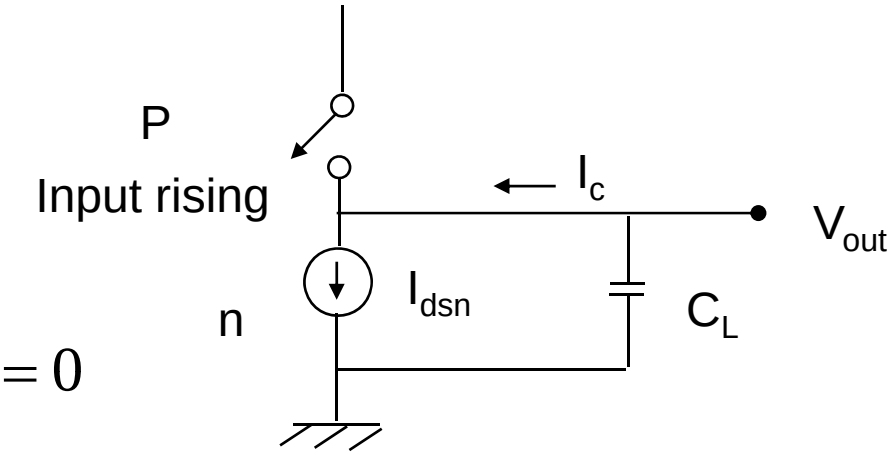
$t_f = t_{f1} + t_{f2}$ ⟶ V_{out} drops from $V_{dd} - V_t$ to $0.1\ V_{DD}$

V_{out} drops from $0.9V_{dd}$ to $V_{dd} - V_t$



Fall Time for Saturated Region

Saturated, $V_{out} \geq V_{DD} - V_{tn}$



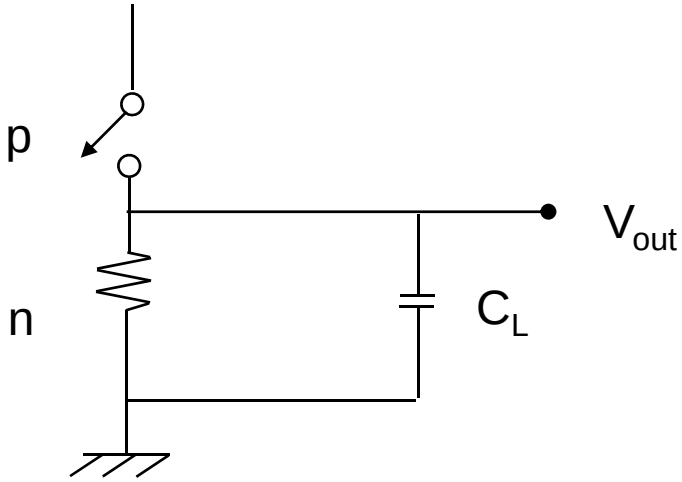
$$C_L \frac{dV_{out}}{dt} + \frac{\beta_n}{2}(V_{DD} - V_{tn})^2 = 0$$

Integrating from $t = t_1$ (corresponding to $V_{out} = 0.9\ V_{DD}$) to $t = t_2$ (corresponding to $V_{out} = (V_{DD} - V_{tn})$)

$$t_{f1} = 2\frac{C_L}{\beta_n (V_{DD} - V_{tn})^2} \int_{V_{DD}-V_{tn}}^{0.9V_{DD}} dV_{out}$$

$$= \frac{2C_L(V_{tn} - 0.1V_{DD})}{\beta_n (V_{DD} - V_{tn})^2}$$

Fall Time for Non-Saturated Region



Non-saturated : $0 \le V_{out} \le V_{DD} - V_{tn}$

$$C_L \frac{dV_{out}}{dt} + \beta_n [(V_{DD} - V_{tn}).V_{out} - \frac{V^2_{out}}{2}] = 0$$

$$t_{f2} = \frac{C_L}{\beta_n (V_{DD} - V_{tn})} \int_{V_{DD}-V_{tn}}^{0.1V_{DD}} \frac{dV_{out}}{\frac{V^2_{out}}{2(V_{DD} - V_{tn})} - V_{out}}$$

Fall Time for Non-Saturated Region

$$t_{f2} = \frac{C_L}{\beta_n (V_{DD} - V_{tn})} \int_{V_{DD}-V_{tn}}^{0.1V_{DD}} \frac{dV_{out}}{\frac{V^2_{out}}{2(V_{DD} - V_{tn})} - V_{out}}$$

$$= \frac{C_L}{\beta_n (V_{DD} - V_{tn})} \ln\left(\frac{19V_{DD} - 20V_{tn}}{V_{DD}}\right)$$

$$= \frac{C_L}{\beta_n V_{DD} (1 - n)} \ln(19 - 20n)$$

where n = $\frac{V_{tn}}{V_{DD}}$

Fall Time Computation

$$t_f = t_{f1} + t_{f2}$$

$$= 2\frac{C_L}{\beta_n V_{DD}(1-n)}\left[\frac{(n-0.1)}{(1-n)} + \frac{1}{2}\ln(19-20n)\right]$$

$$\boxed{t_f \approx k\frac{C_L}{\beta_n V_{DD}}}$$

$k = 3 \sim 4$ for $V_{DD} = 3 \sim 5V$ and $V_{tn} = 0.5 \sim 1V$

Rise Time

$$t_r = 2\frac{C_L}{\beta_p V_{DD}(1-p)}\left[\frac{(p-0.1)}{(1-p)} + \frac{1}{2}\ln(19-20p)\right]$$

with $p = \frac{|V_{tp}|}{V_{DD}}$

$$t_r \approx k\frac{C_L}{\beta_p V_{DD}}$$

For equally sized n- and p transistors

$$\beta_n \approx 2\beta_p$$

$$\boxed{t_f \approx \frac{t_r}{2}}$$

Sizing for Identical Rise/Fall Time

For same t_f and t_r

$$\frac{\beta_n}{\beta_p} = 1$$

Increase the width of p-device to

$$W_p \approx 2 - 3W_n$$

Delay Time: First Order Approximation

- Gate delay is dominated by the output rise and fall time

$$t_{dr} = \frac{t_r}{2}$$

$$t_{df} = \frac{t_f}{2}$$

General Delay Time Computation

- Similar to the computation of rise/fall times
  - Saturation region from t = t_1 (corresponding to $V_{out} = V_{DD}$) to t = t_2 (corresponding to $V_{out} = (V_{DD} - V_{tn})$)
  - Linear region from t = t_2 (corresponding to $V_{out} = (V_{DD} - V_{tn})$) to t = t_3

$$t_2 - t_1 = 2\frac{C_L}{\beta_n (V_{DD} - V_{tn})^2} \int_{V_{DD}-V_{tn}}^{V_{DD}} dV_{out}$$

$$= \frac{2C_L (V_{tn})}{\beta_n (V_{DD} - V_{tn})^2}$$

Delay Time Computation

$$t_3 - t_2 = \frac{C_L}{\beta_n (V_{DD} - V_{tn})} \int_{V_{DD}-V_{tn}}^{V_{out}} \frac{dV_{out}^{'}}{\frac{V_{out}^{'2}}{2(V_{DD} - V_{tn})} - V_{out}^{'}}$$

$$= \frac{C_L}{\beta_n (V_{DD} - V_{tn})} \ln\left(\frac{2V_{DD} - 2V_{tn} - V_{out}}{V_{out}}\right)$$

$$= \frac{C_L}{\beta_n V_{DD} (1-n)} \ln\left(\frac{2(1-n) - V_O}{V_O}\right)$$

where $n = \frac{V_{tn}}{V_{DD}}$, $V_O = \frac{V_{out}}{V_{DD}}$

Delay Time

$$t_{Dn} = t_3 - t_1 = A_n \frac{C_L}{\beta_n V_{DD}}$$

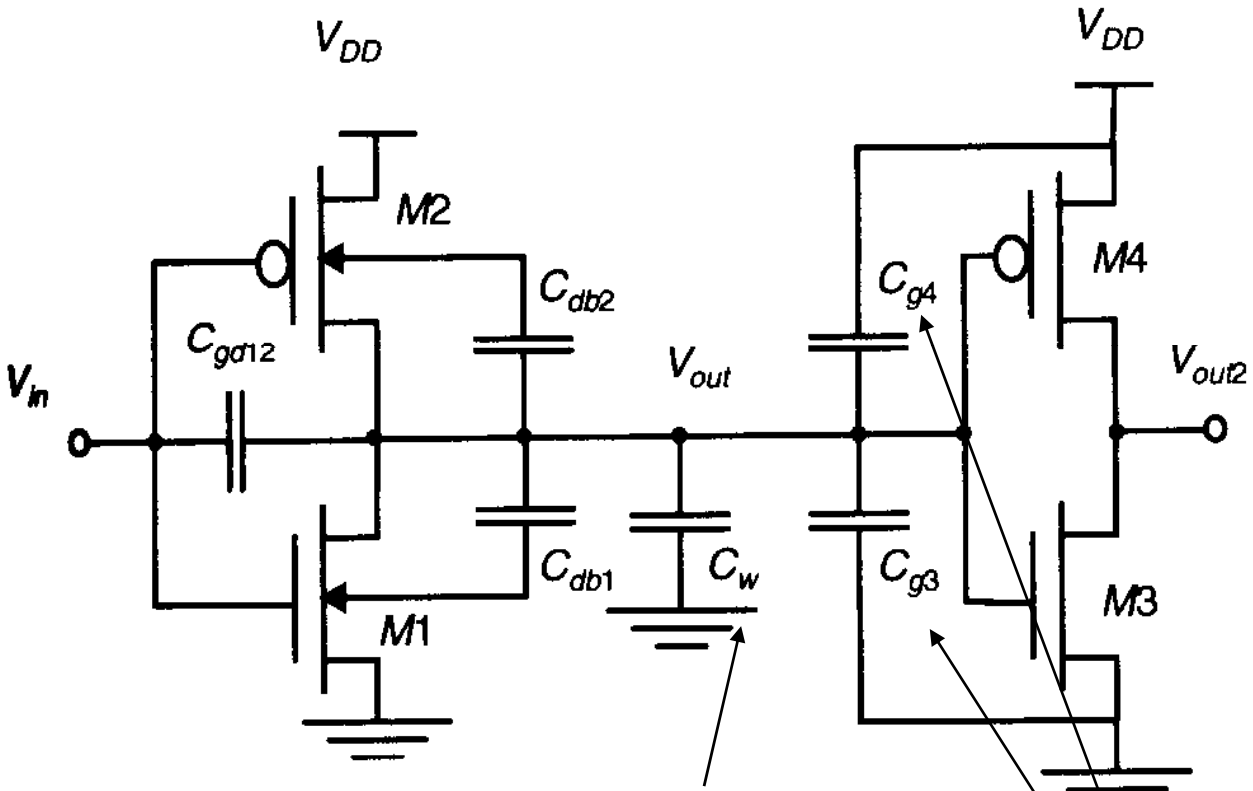
Delay $\propto C_L$ (optimize C_L to decrease delay)

$\propto \frac{1}{V_{DD}}$ (decrease V_{DD} increases delay)

$\propto \frac{1}{\beta}$ (if W $\uparrow$ or L $\downarrow$, delay decreases)

Three major parameters for optimizing speed of CMOS

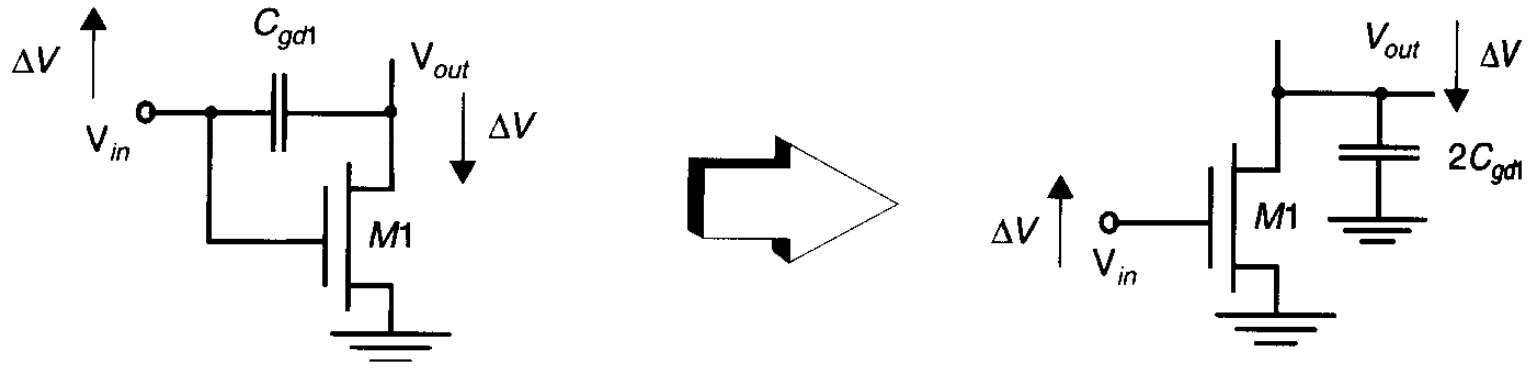
Components of C_L



C_w = wiring capacitance

C_g = gate capacitance = $C_{ox}WL$

Miller Effect



- Effective voltage change over the gate-drain capacitor is actually twice the output voltage swing
- Contribution of gate-drain capacitor should be counted twice

Junction Capacitance

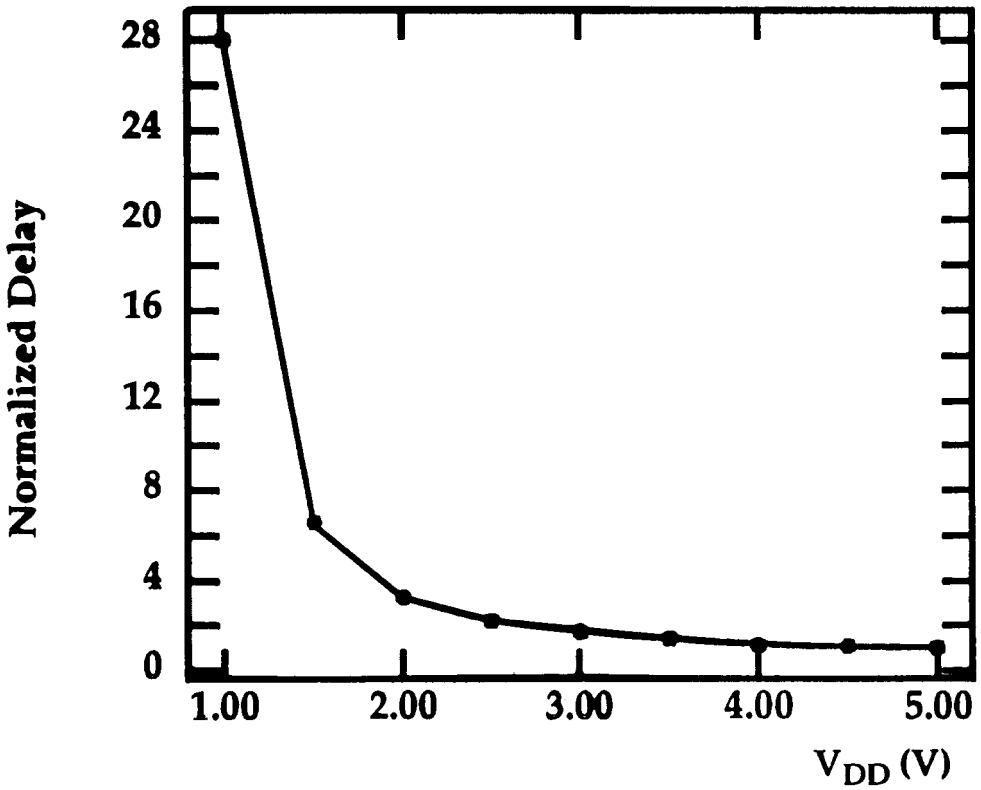
- Non-linear capacitor modeled by linear capacitor with the same change in charge for the voltage range of interest

$$C_{eq} = K_{eq} C_{j0}$$

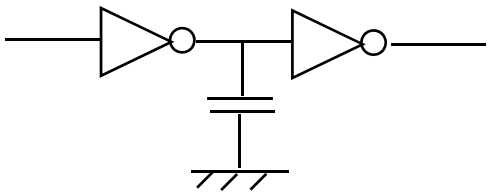
$$K_{eq} = \frac{-\phi_0^m}{(V_{high} - V_{low})(1-m)} \left[(\phi_0 - V_{high})^{1-m} - (\phi_0 - V_{low})^{1-m} \right]$$

- Linearize over the interval {5V, 2.5V} for the high-to-low transition and {0, 2.5V} for the low-to-high transition
- Correspond to {V_{high}=-5V, V_{low}=-2.5V} and {V_{high}=0, V_{low}=-2.5V} for NMOS

Delay in function of V_{DD}



Sizing of Inverter Loaded by an Identical Gate



Load cap. of first gate:

$$C_L = (C_{dp1} + C_{dn1}) + (C_{gp2} + C_{gn2}) + C_W$$

where

C_{dp1}, C_{dn1} $\rightarrow$ diffusion capacitance of first gate

C_{gp2}, C_{gn2} $\rightarrow$ gate capacitance of second gate

C_w $\rightarrow$ wiring capacitance

If PMOS devices are α times larger than the NMOS ones,

$$\alpha = (W/L)_p \Big/ (W/L)_n$$

all transistor capacitances will scale in approximately the same way

Sizing of Inverter

$$C_{dp1} \approx \alpha C_{dn1}$$

$$C_{gp2} \approx \alpha C_{gn2}$$

$$\therefore C_L = (1+\alpha)(C_{dn1} + C_{gn2}) + C_w$$

$$t_p = \frac{t_r + t_f}{2} = \frac{C_L}{2V_{DD}}\left(\frac{A_n}{\beta_n} + \frac{A_p}{\beta_p}\right)$$

$$= \frac{C_L}{2V_{DD}.\beta_n}\left(A_n + \frac{A_p\beta_n}{\beta_p}\right)$$

$$= \frac{C_L}{2V_{DD}.\beta_n}\left(A_n + \frac{A_p\mu_n}{\mu_p}.\frac{(W/L)_n}{(W/L)_p}\right)$$

$$= \frac{C_L}{2V_{DD}.\beta_n}\left(A_n + \frac{A_p\mu_n}{\mu_p\alpha}\right)$$

Sizing of Inverter

$$t_p = \frac{C_L}{2V_{DD}.\beta_n}(A_n + \frac{A_p\mu_n}{\mu_p\alpha})$$

$$= \frac{(1+\alpha)(C_{dn1} + C_{gn2}) + C_W}{2V_{DD}.\beta_n}(A_n + \frac{A_p\mu_n}{\mu_p.\alpha})$$

$$\text{Let } \frac{\partial t_p}{\partial \alpha} = 0 \text{ to get optimal } \alpha$$

$$\alpha_{opt} = \sqrt{\frac{\mu_n}{\mu_p}\frac{A_p}{A_n}(1 + \frac{C_W}{C_{dn1} + C_{gn2}})}$$

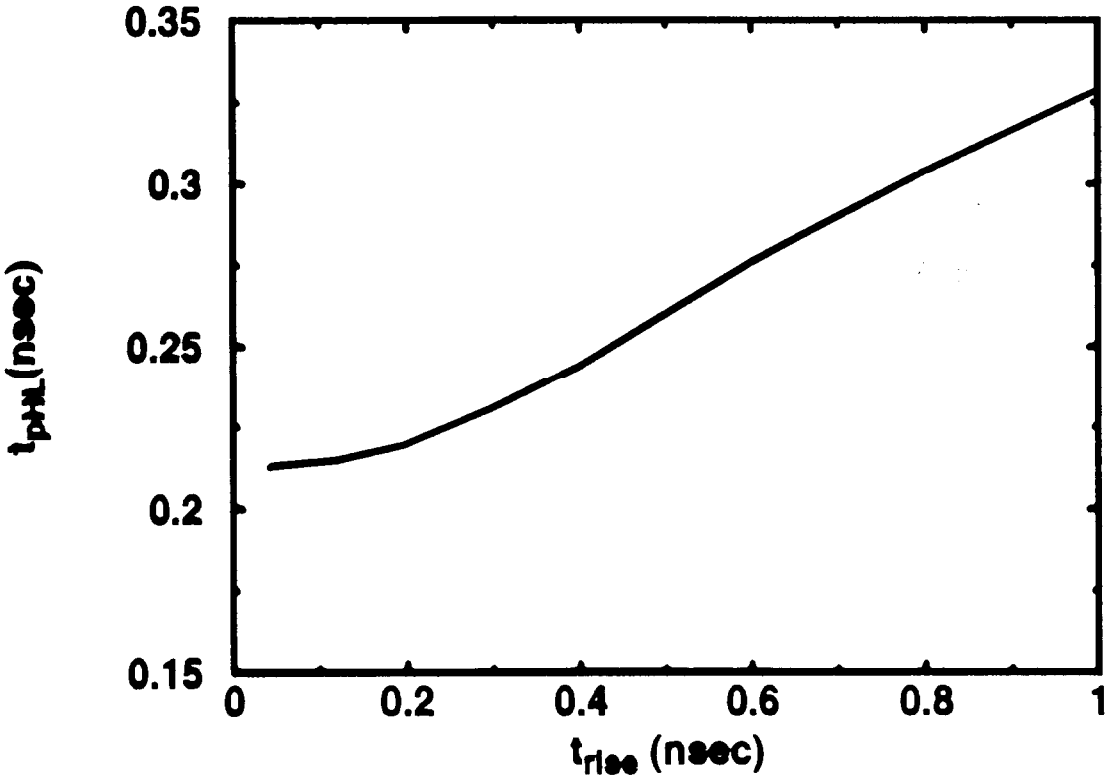
If $C_W << C_{dn1} + C_{gn2}$, $A_p = A_n$

$$\alpha_{opt} \approx \sqrt{\frac{\mu_n}{\mu_p}} \approx 1.73 \longrightarrow$$

Contrast to 3 which is normally used in the non-cascaded case

Impact of Rise Time on Delay

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



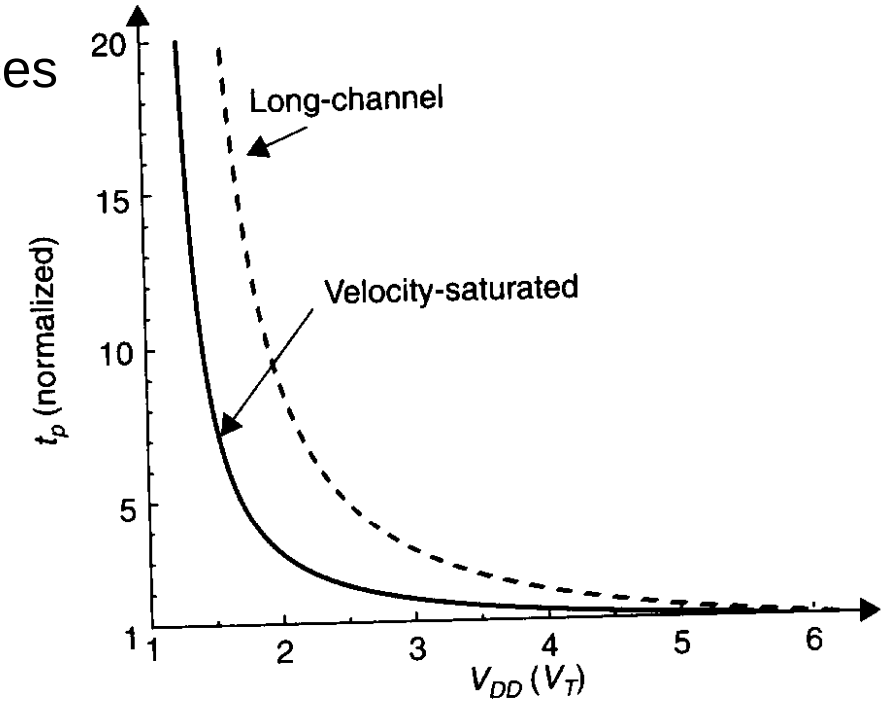
Minimum-size inverter with fanout of a single gate

Velocity Saturation

- Under long channel model, saturation current $\propto V_{DD}^2$
- In small-geometry devices, this no longer holds: $I_{av} \propto V_{DD}$
- Therefore, for $V_{DD} >> V_T$ we have,

$$t_p \approx \frac{C_L}{2}\left(\frac{1}{k_p}+\frac{1}{k_n}\right) \qquad k_{n,p} = \kappa v_{SAT} C_{ox} W_{n,p}$$

- Running velocity saturated devices at high V_{DD} is not beneficial
- Lowering V_{DD} below $2V_T$ sharply increases delay



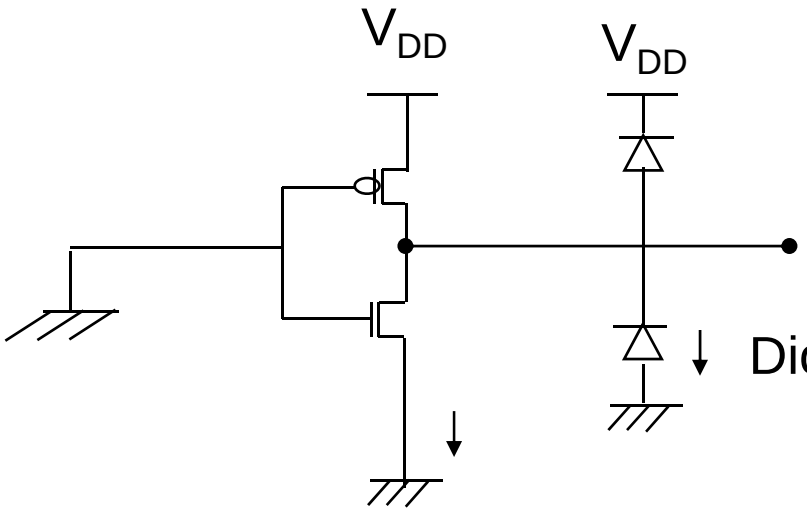
Source/Drain Resistance

- In small-geometry devices, source and drain resistance affects switching currents
  - Source of the transistor is no longer grounded, body effect increases threshold voltage
  - V_{gs} is also reduced
  - Current is reduced

Power Consumption

- Static Power
  - Leakage current
  - Sub-threshold conductance
- Dynamic Power
  - Capacitive Power due to charging/discharging of capacitive load
  - Short-circuit power due to direct path currents when there is a temporary connection between power and ground

Static Power Consumption



$$I_O = i_s(e^{Vq/kT} - 1)$$

Sub-threshold current

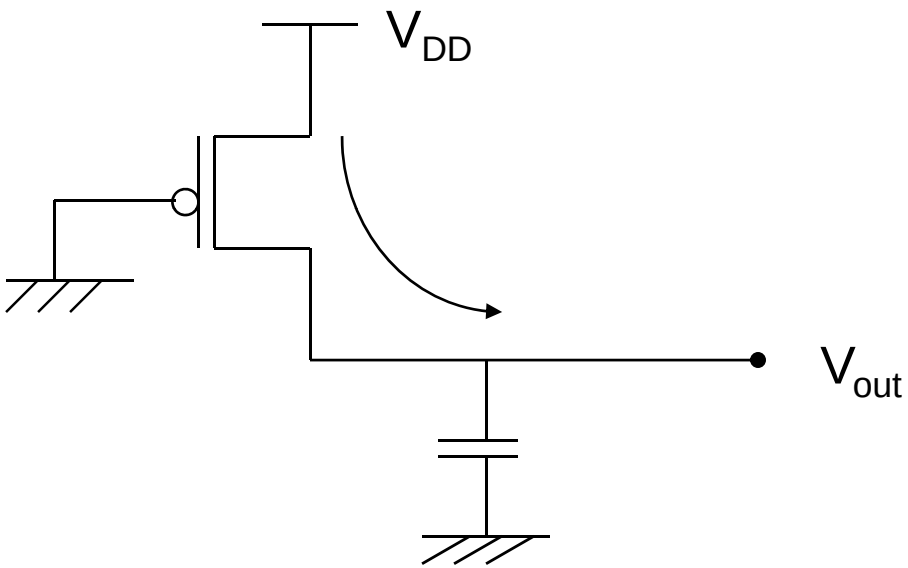
$$I_D = K \cdot e^{(V_{gs} - V_t)q/nkT}(1 - e^{V_{ds}q/kT})$$

$P_{static} = I_{leakage} . V_{DD}$

Static Consumption

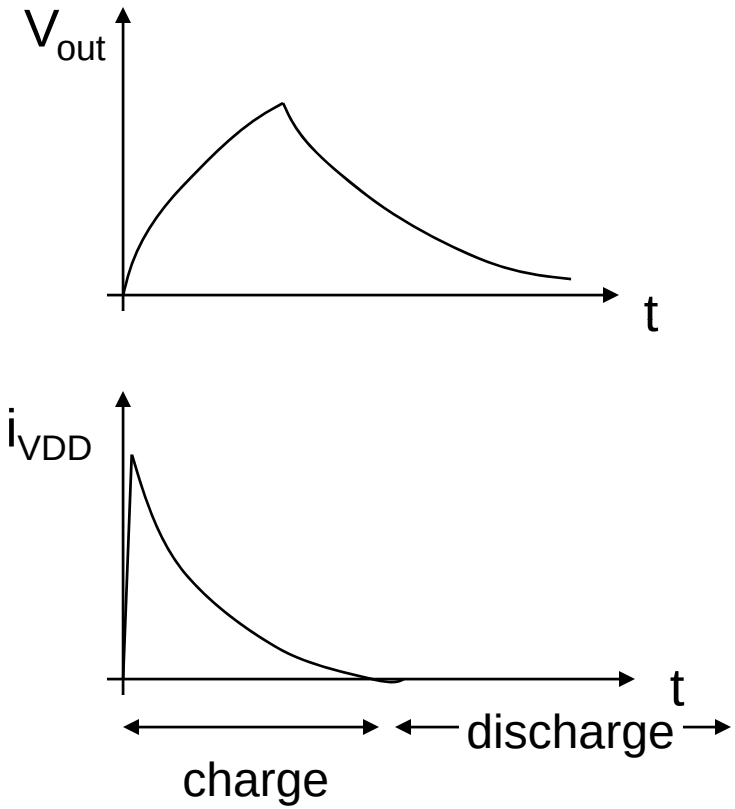
- Leakage current through the reverse biased diode junctions
- For typical devices it is between 0.1nA - 0.5nA at room temperature
- For a die with 1 million devices operated at 5 V, this results in 0.5mW power consumption $\rightarrow$ not much
- Junction leakage current is caused by thermally generated carriers -> therefore is a strong function of temperature
- More important is sub-threshold leakage when threshold voltage is close to 0

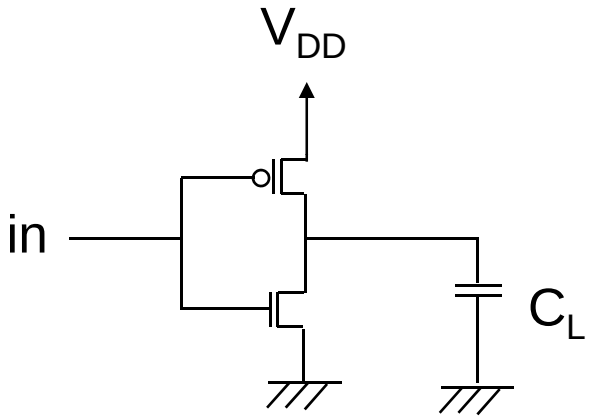
Dynamic Consumption due to C_L



- low-to-high transition
- Assume 0 rise and fall times

Dynamic Power due to C_L





Define:

E_{VDD} : energy taken from supply during a transition

E_C: energy stored on capacitor at the end of transition

Energy Consumed and Stored

$$E_{V_{DD}} = \int_0^\infty i_{VDD}(t) V_{DD} dt = V_{DD} \int_0^\infty C_L \frac{dV_{out}}{dt}.dt$$

$$= C_L V_{DD} \int_0^{V_{DD}} dV_{out}$$

$$= C_L V_{DD}{}^2 (= Q V_{DD})$$

$$E_C = \int_0^\infty i_{VDD}(t) V_{out} dt = \int_0^\infty C_L \frac{dV_{out}}{dt} V_{out} dt$$

$$= C_L V_{DD} \int_0^{V_{DD}} V_{out} dV_{out}$$

$$= \frac{C_L . V_{DD}{}^2}{2}$$

Half the energy is stored in Capacitor ! Other half is dissipated in the PMOS transistor !!

For each switching cycle (L $\rightarrow$ H & H $\rightarrow$ L), amount of energy dissipated in $C_L . V_{DD}{}^2$

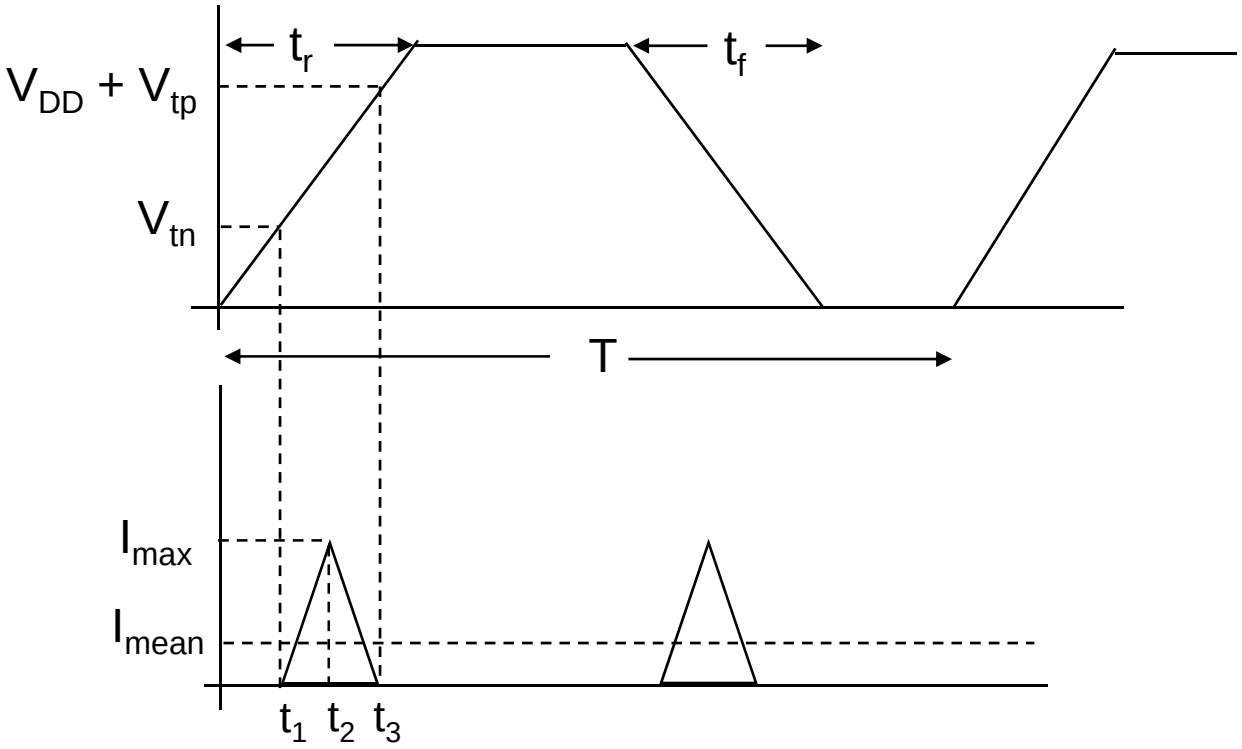
$$P_{dynamic} = C_L.V_{DD}^2.f$$

- Example
  - 1.2μ CMOS chip
  - 100 MHz clock rate
  - Average load capacitance of 30 fF/gate
  - 5V power supply
- Power consumption/gate = 75 μW
- Design with 200,000 gates: 15W !
- Pessimistic evaluation: not all gates switch at the full rate
- Have to consider the activity factor α: Effective switching capacitance = αC_L
- Reducing V_{DD} has a quadratic effect on $P_{dynamic}$

Direct Path Current

- inputs have finite rise and fall times
- Direct current path from V_{DD} to GND while PMOS and NMOS are ON simultaneously for a short period

$$P_{sc} = I_{mean}.V_{DD}$$



Symmetrical Inverter Without Load

$$I_{mean} = 2\left[\frac{1}{T}\int_{t_1}^{t_2} I(t)dt + \frac{1}{T}\int_{t_2}^{t_3} I(t)dt\right]$$

If $V_{tn} = -V_{tp}=V_T$ and $\beta_n = \beta_p = \beta$
and that the behavior around t_2 is symmetrical

$$I_{mean} = 2\times\frac{2}{T}\int_{t_1}^{t_2}\frac{\beta}{2}(V_{in}(t)-V_t)^2\,dt$$

with $$V_{in}(t) = \frac{V_{DD}}{t_r}t$$

$$t_1 = \frac{V_t}{V_{DD}}.t_r$$

$$t_2 = \frac{t_r}{2}$$

$$t_r = t_f = t_{rf}$$

Symmetrical Inverter Without Load

$$I_{mean} = \frac{2\beta}{T}\int_{t_{rf}V_T/V_{DD}}^{t_{rf}/2}\left(\frac{V_{DD}}{t_{rf}}\cdot t - V_t\right)^2 dt$$

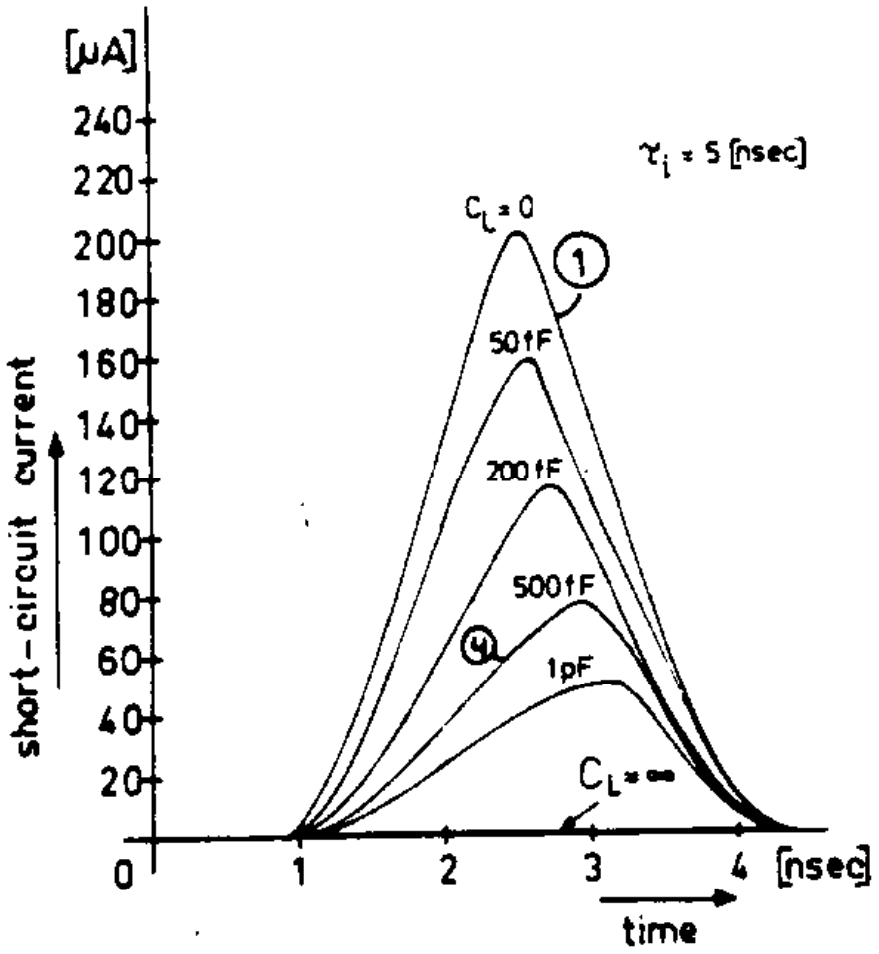
$$= \frac{2\beta}{T}\left[\frac{t_{rf}}{3V_{DD}}\left(\frac{V_{DD}}{t_{rf}}\cdot t - V_t\right)^3\right]_{t_{rf}V_T/V_{DD}}^{t_{rf}/2}$$

$$= \frac{2t_{rf}}{3T}\cdot\frac{\beta}{V_{DD}}\left(\frac{V_{DD}}{2} - V_t\right)^3$$

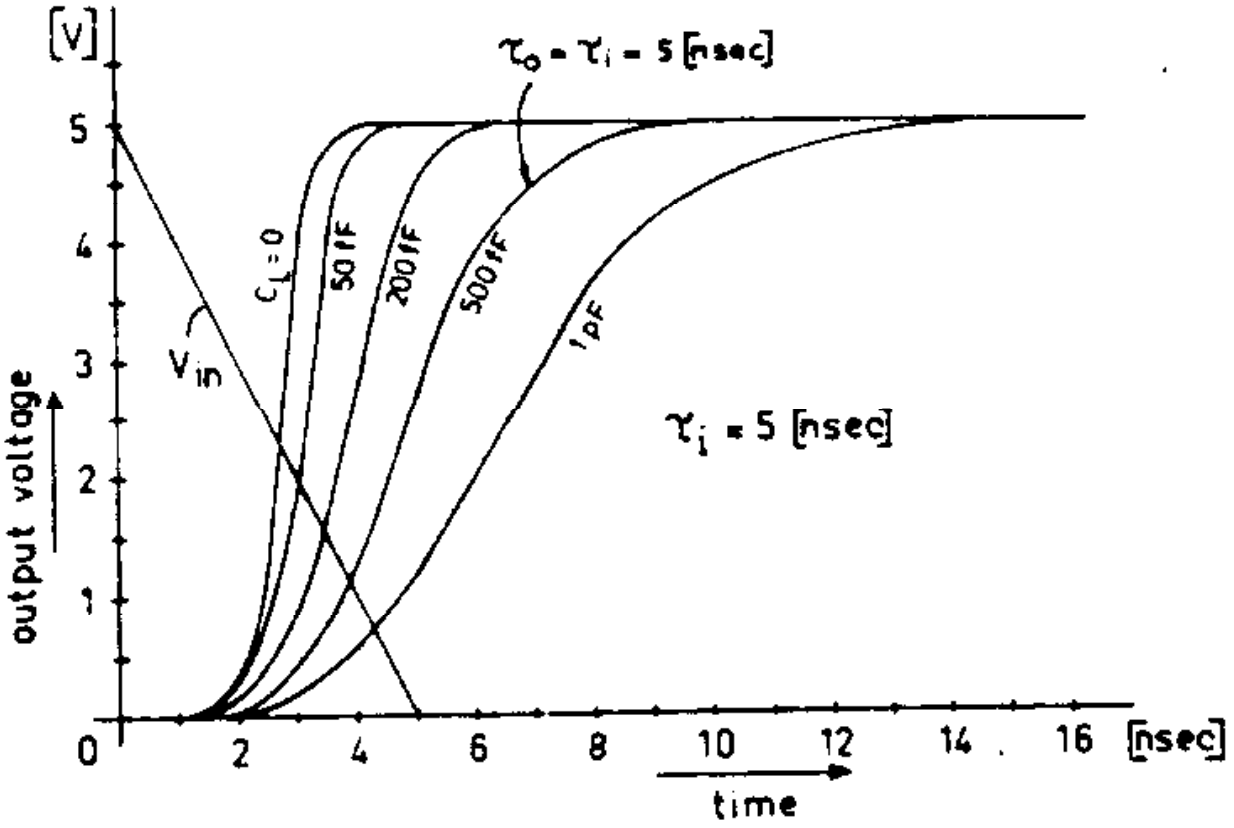
$$= \frac{t_{rf}}{12T}\cdot\frac{\beta}{V_{DD}}\left(V_{DD} - 2V_t\right)^3$$

$$P_{sc} = \frac{\beta}{12}\left(V_{DD} - 2V_t\right)^3\frac{t_{rf}}{T}$$

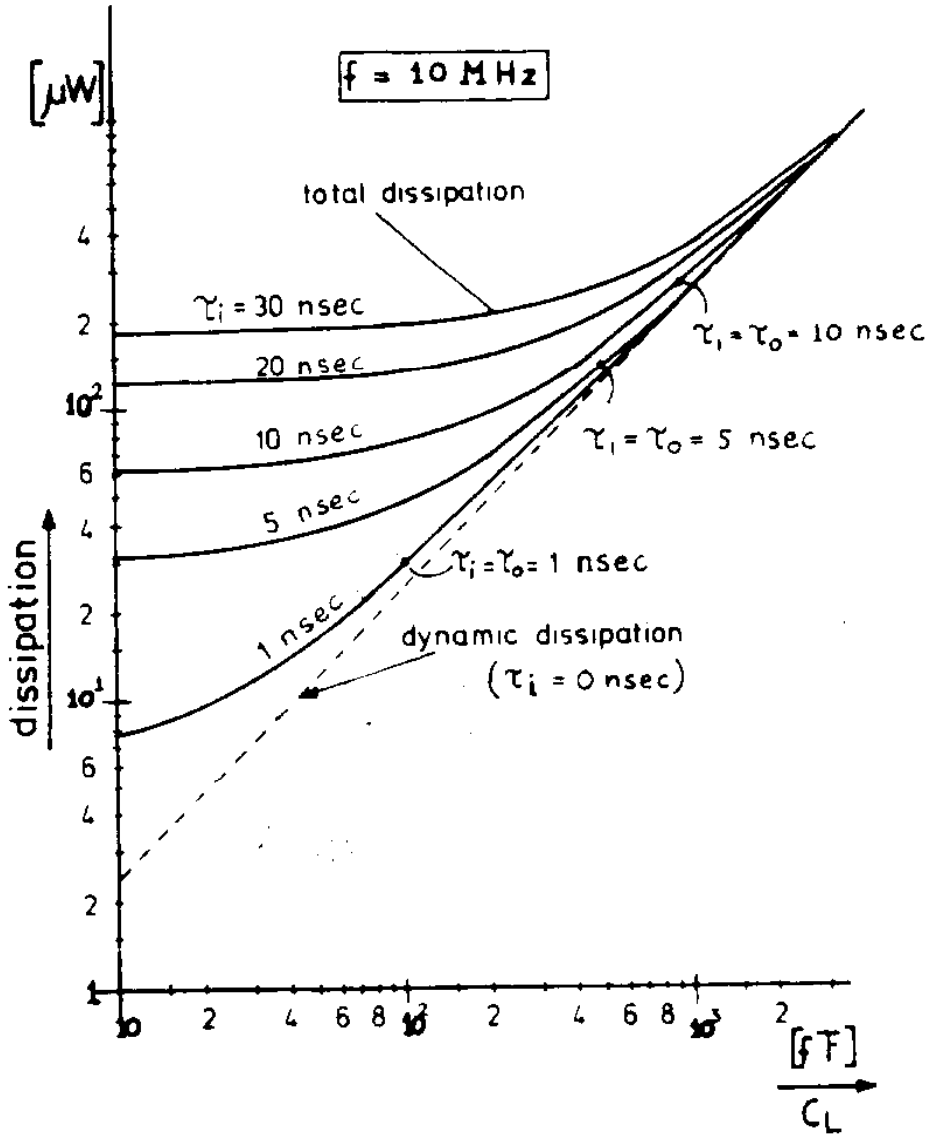
Short Circuit Current with Loads



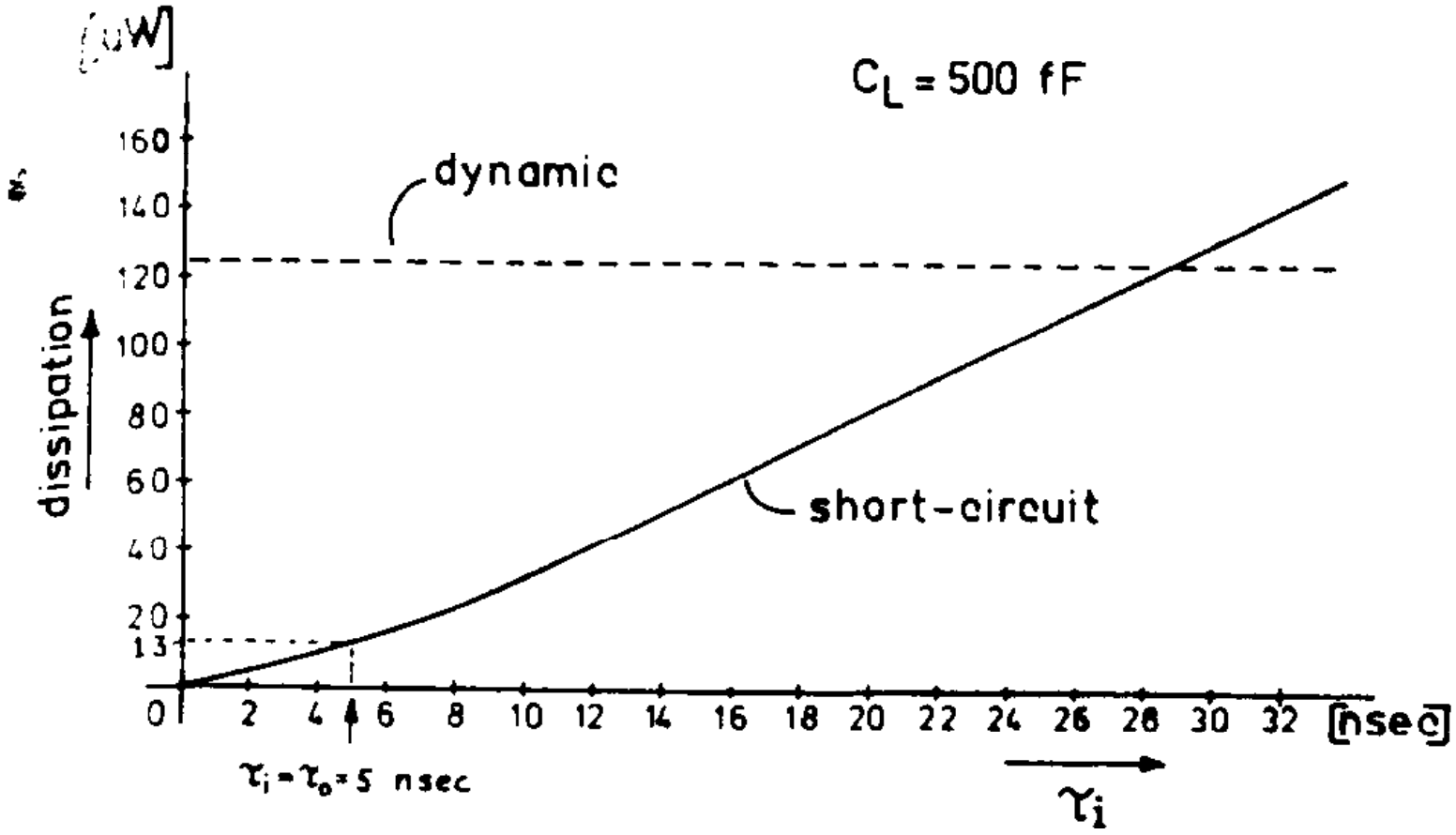
Output Transitions under Different Loads



C_L Power vs. SC Power under Different Loads



C_L Power vs. SC Power under Different Inputs



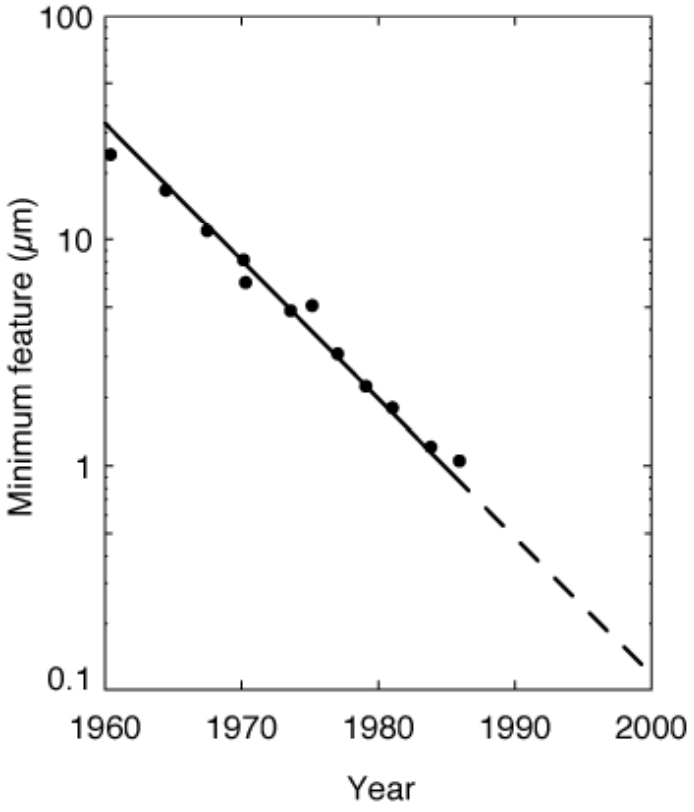
Impact of Load Capacitance on SC Current

- Large capacitance
  - Fast input transition, slow output transition
  - Input moves through the transient region before output begins to change
  - Short-circuit current close to zero
- Small capacitance
  - Relatively slower input transition, fast output transition
  - Both devices in saturation during most of the transition
  - Maximum short-circuit current
- [Veendrick84]: rise/fall times of all signals should be kept constant within a range to keep SC power minimal, 10%~20% of total dynamic power

Technology Evolution

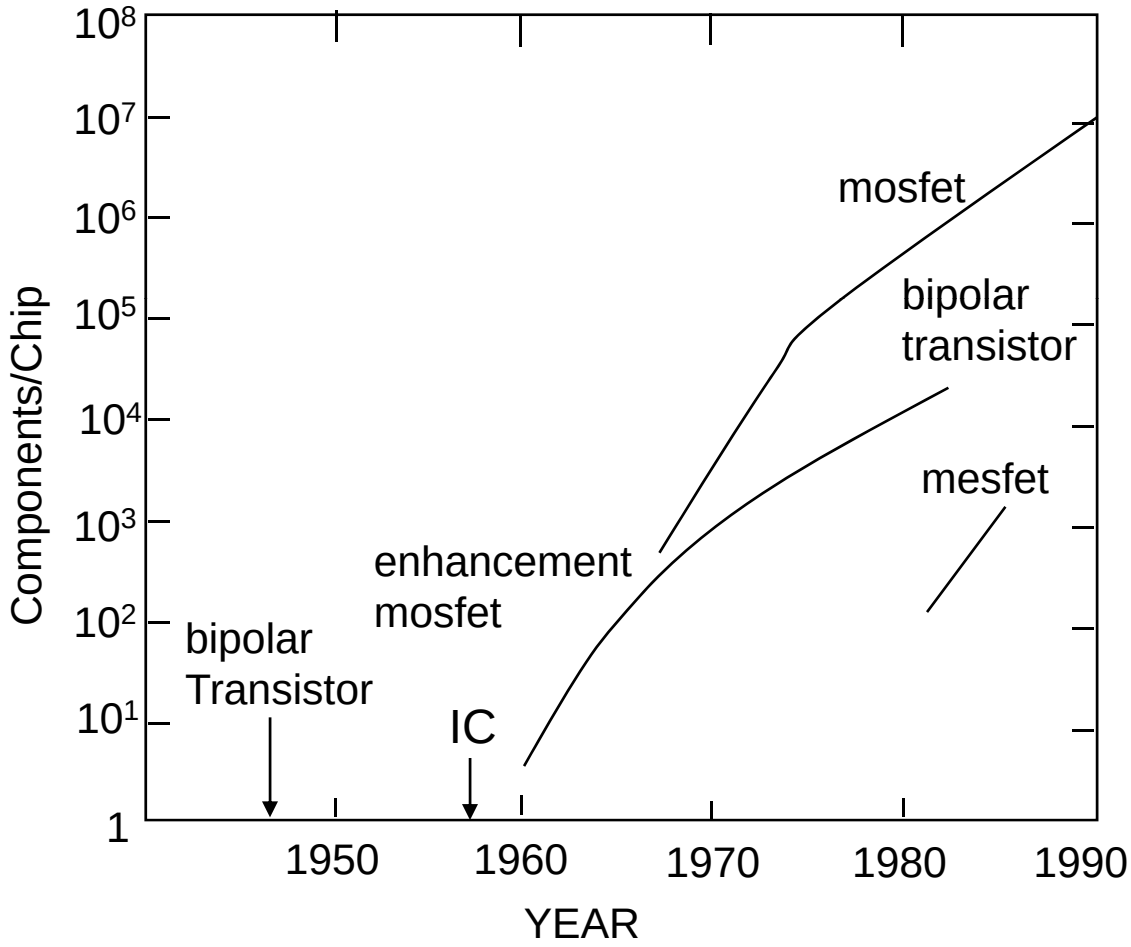
Year of Introduction	1994	1997	2000	2003	2006	2009
Channel length (μm)	0.4	0.3	0.25	0.18	0.13	0.1
Gate oxide (nm)	12	7	6	4.5	4	4
V_{DD} (V)	3.3	2.2	2.2	1.5	1.5	1.5
V_T (V)	0.7	0.7	0.7	0.6	0.6	0.6
NMOS I_{Dsat} (mA/μm) (@ $V_{GS} = V_{DD}$)	0.35	0.27	0.31	0.21	0.29	0.33
PMOS I_{Dsat} (mA/μm) (@ $V_{GS}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

Technology Scaling (1)



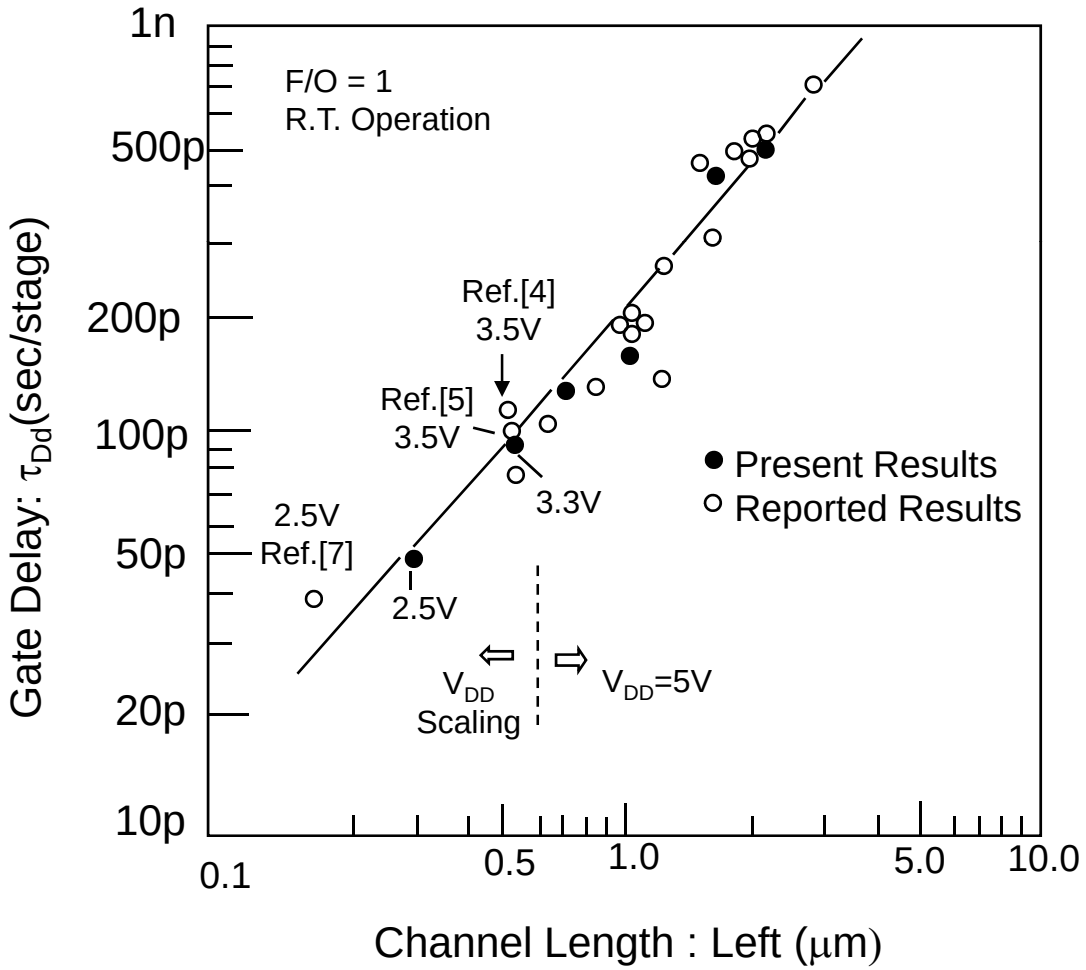
Minimum Feature Size

Technology Scaling



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

Scaling Relationships for Long channel Devices

Parameter	Relation	Full Scaling	General Scaling	Fixed Voltage Scaling
W, L, t_{ox}		1/S	1/S	1/S
V_{DD}, V_T		1/S	1/U	1
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Area/Device	WL	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_L	$C_{ox}WL$	1/S	1/S	1/S
k_n, k_p	$C_{ox}W/L$	S	S	S
I_{av}	$k_{n,p} V^2$	1/S	S/U^2	S
t_p (intrinsic)	$C_L V / I_{av}$	1/S	U/S^2	$1/S^2$
P_{av}	$C_L V^2 / t_p$	$1/S^2$	S/U^3	S
PDP	$C_L V^2$	$1/S^3$	$1/SU^2$	1/S

Scaling of Short Channel Devices

Parameter	Relation	Full Scaling	General Scaling	Fixed Voltage Scaling
I_{av}	$C_{ox}WV$	1/S	1/U	1
J_{av}	I_{av}/Area	S	S^2/U	S^2
t_p (intrinsic)	$C_L V / I_{av}$	1/S	1/S	1/S
P_{av}	$C_L V^2 / t_p$	$1/S^2$	$1/U^2$	1

Homework Problem (due next Thursday)

- Design a static CMOS inverter with 0.4pF load capacitance. Make sure that you have equal rise and fall times. Layout the inverter using the Mentor tools, extract parasitics, and simulate the extracted circuit on HSPICE to make sure that your design conforms to the specification.
- Do the same analysis for a three input NAND gate.