Lecture 4
The CMOS Inverter

Dynamic properties

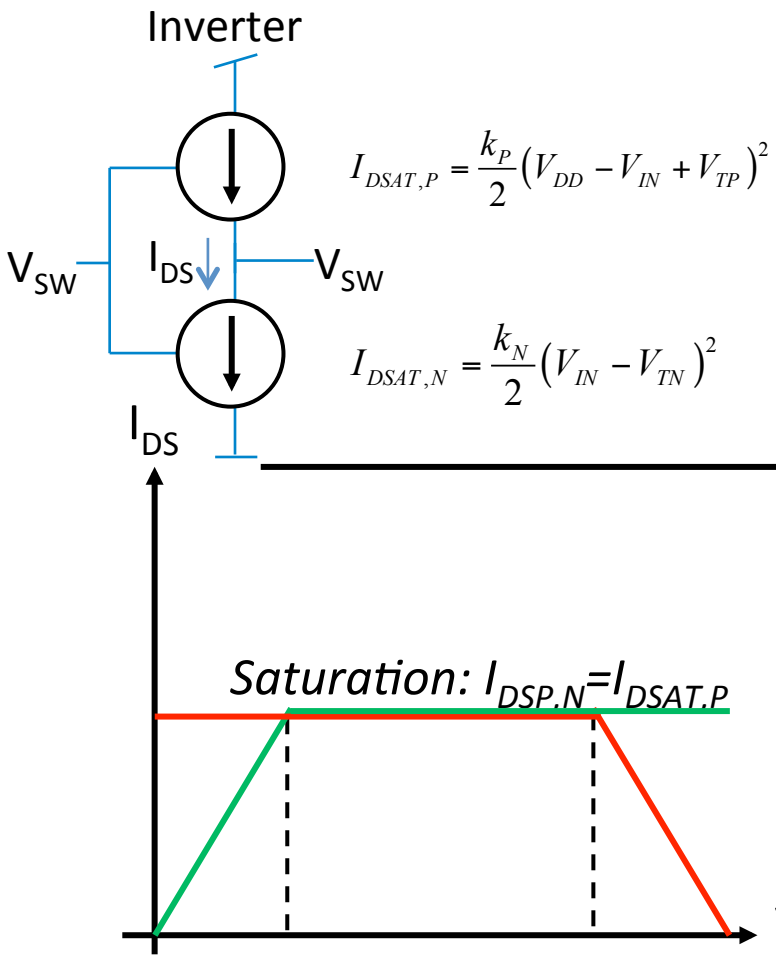
Week 2: The CMOS inverter

- Tuesday
  - Static properties
    - Voltage transfer curves (VTC)
    - Noise margins
  - Exercise (Kjell, 1 hour)
- Thursday
  - Dynamic properties
    - Propagation delay
    - Driving large capacitive loads w. optimal propagation delay
  - Exercise POTW (Victor, 2 hours)
- Friday
  - Prelab 1 deadline 1PM

Last time? Muddy?

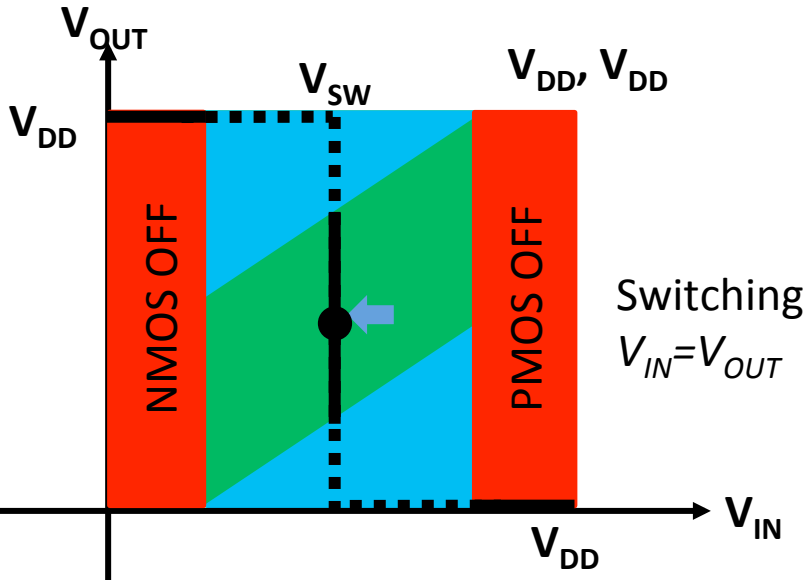
- How switching voltage is calculated
- NOR/NAND gate VTC
- Noise margins
  - How calculate them
  - Are they accurate from simplified models?
  - Where does 1/8 come from in noise margin example?
  - Butterfly diagram
- C_{ox} what is that?

Voltage Transfer Characteristic - VTC



$$I_{DSAT,P} = \frac{k_P}{2}\left(V_{DD} - V_{IN} + V_{TP}\right)^2$$

$$I_{DSAT,N} = \frac{k_N}{2}\left(V_{IN} - V_{TN}\right)^2$$



Switching means:
$V_{IN}=V_{OUT}$

Switching occurs in the **green region** where both MOSFETs are saturated . . .

. . . and saturation currents are equal:

$$\frac{k_N}{2}\left(V_{IN} - V_{TN}\right)^2 = \frac{k_P}{2}\left(V_{DD} - V_{IN} + V_{TP}\right)^2$$

Solving for V_{IN} using $x=k_N/k_P$ yields

$$V_{sw} = \frac{V_{DD} + V_{TP} + \sqrt{x}V_{TN}}{1+\sqrt{x}}$$

Noise margins

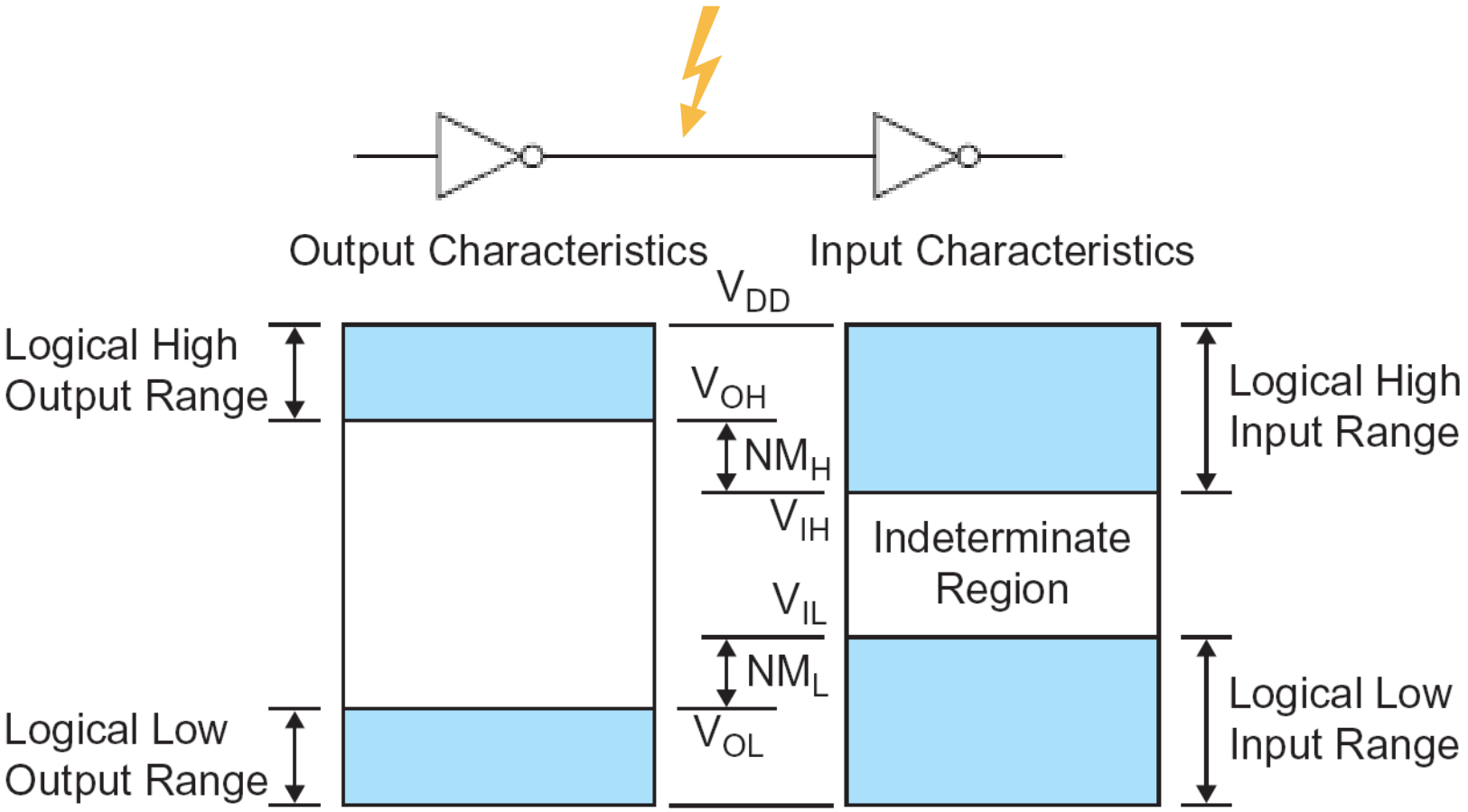


FIGURE 2.29 Noise margin definitions

Noise Margins – an example

Let´s define valid regions from points where slope $A_V = -1$!

These points yields numbers for ($V_{OH,min}$, $V_{IL,max}$) and ($V_{OL,max}$, $V_{IH,min}$) so that NMH and NML can be calculated!

For x=1, V_{TN}=0.28 V and V_{TP}=-0.28 V we have V_{SW}=0.28+0.64/2=0.60 V and ΔV=0.64 V

Formulas can be derived (for x=1):

$V_{OH,min}=V_{DD}-\Delta V/8=1.12$ V

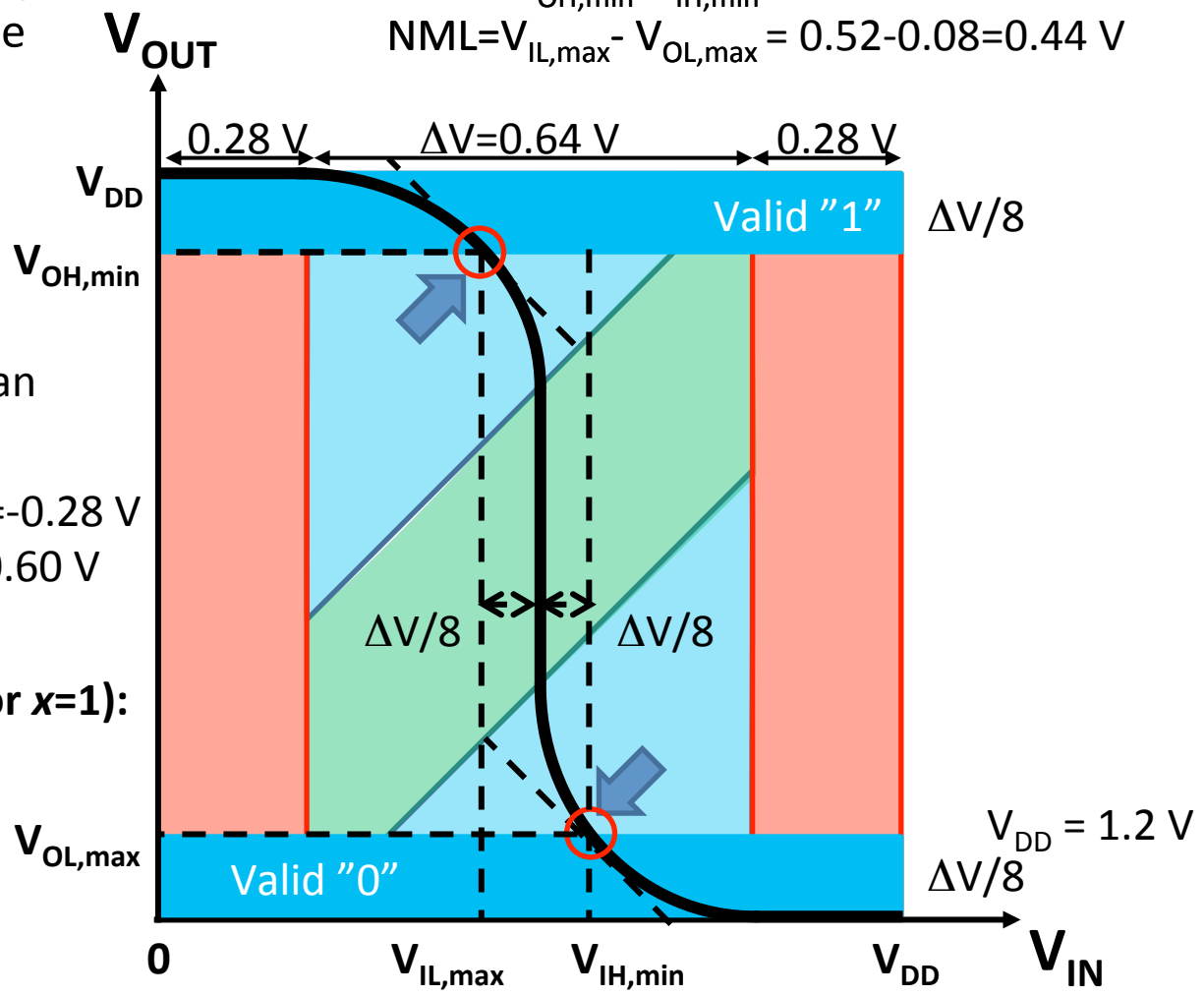
$V_{OL,max}=\Delta V/8=80$ mV

$V_{IL,max}=V_{SW}-\Delta V/8=0.52$ V

$V_{IH,min}=V_{SW}+\Delta V/8=0.68$ V

NMH=$V_{OH,min}$-$V_{IH,min}$1.12-0.68=0.44 V

NML=$V_{IL,max}$- $V_{OL,max}$ = 0.52-0.08=0.44 V

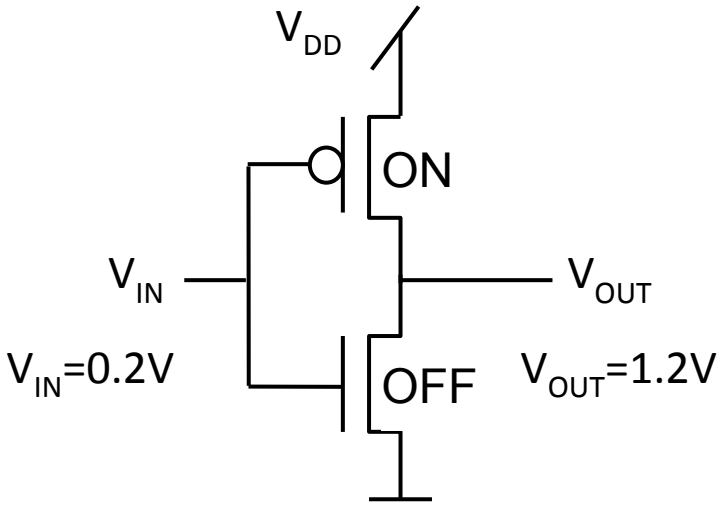


Quick question

- Have you tried solving prelab 3 yet?

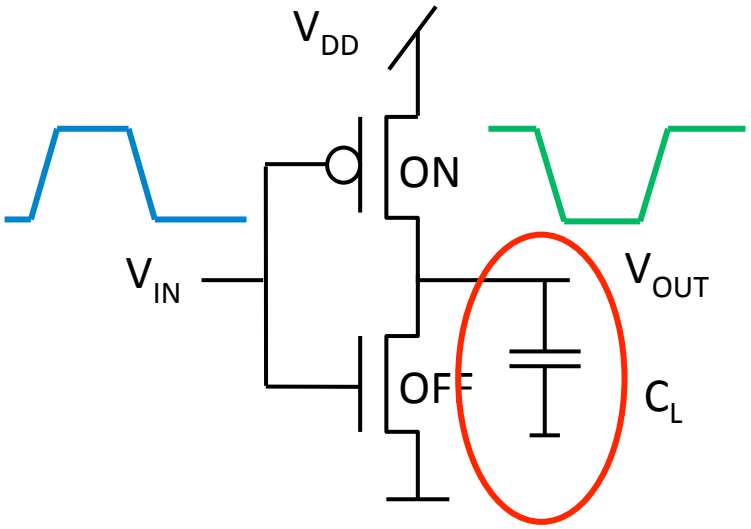
Static vs dynamic

Last time: **static** (DC) behavior
Input: DC voltage,
Output: DC voltage



In circuit simulation: **DC** analysis

This time**: dynamic** behavior
That is, with time
Input: voltage waveform
Output voltage waveform

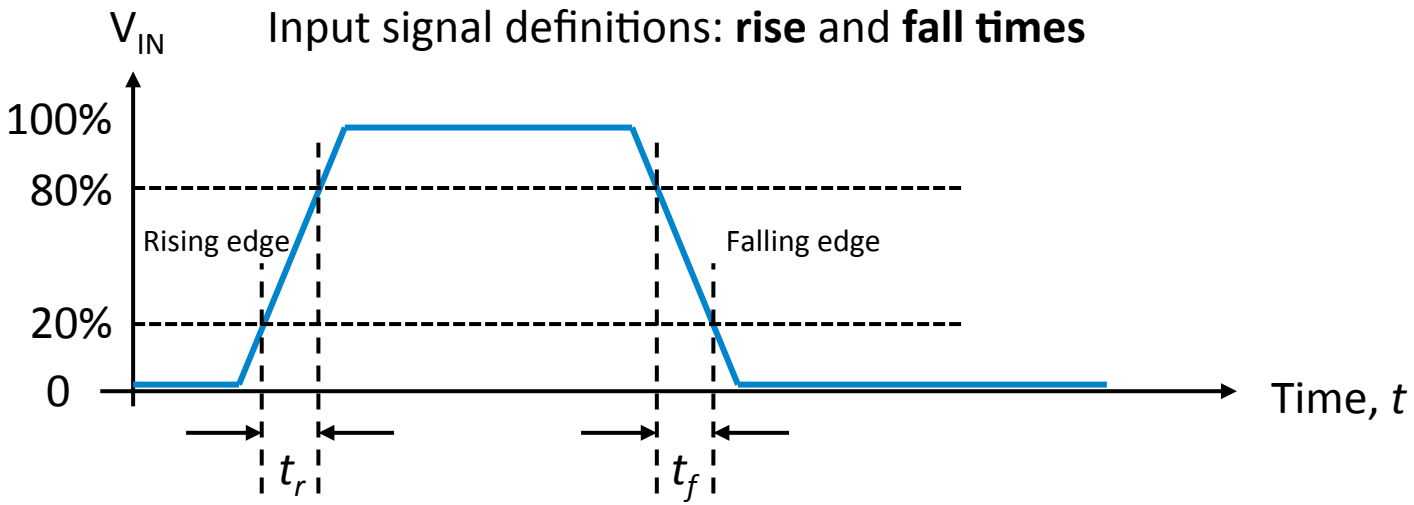


In circuit simulation: **transient** analysis

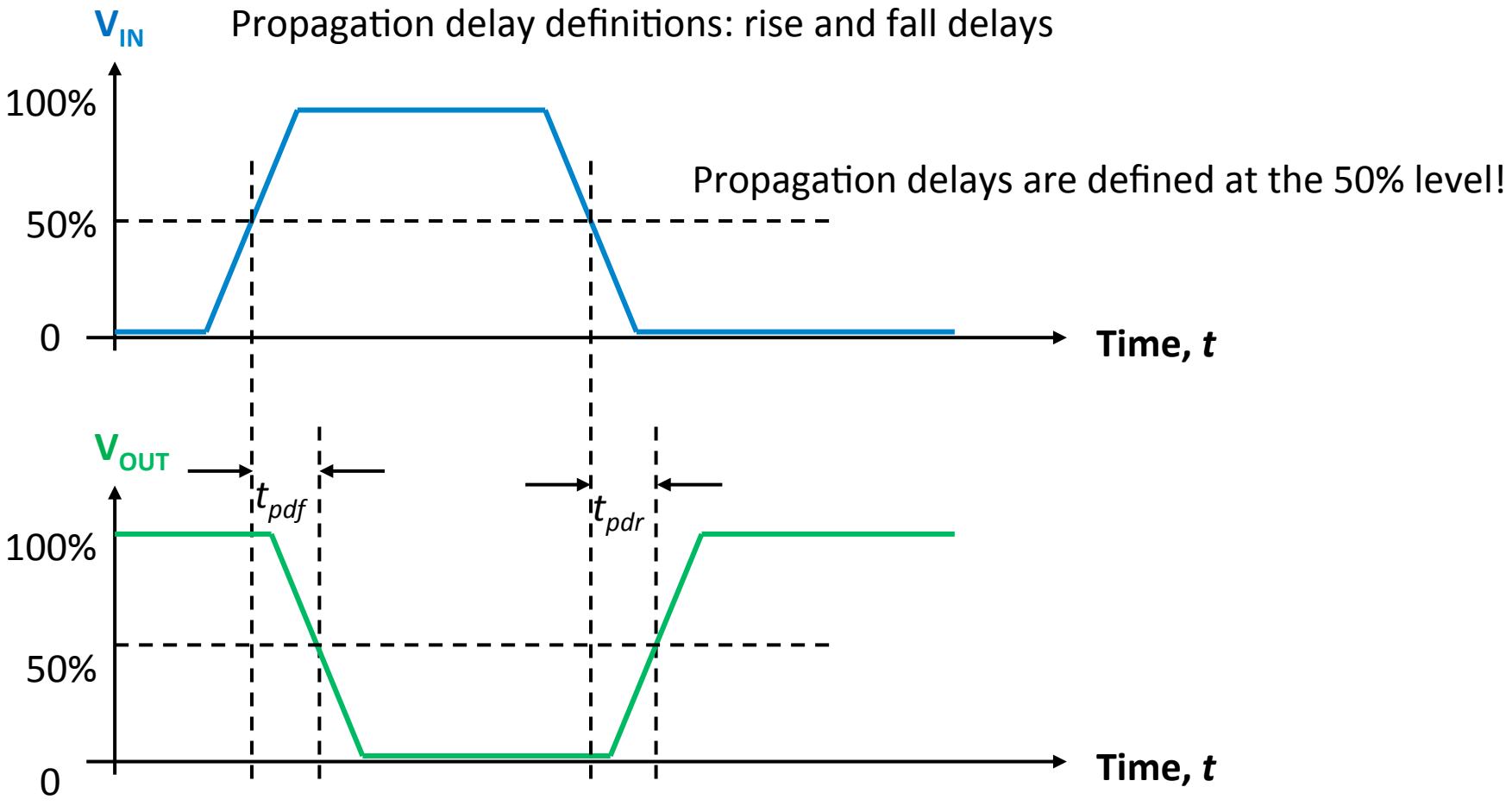
Outline

- Definitions
  - Rise time and fall time
  - Propagation delay: Rise delay and fall delay
- Propagation delay estimation
  - Step response model
    - Charging and discharging the load capacitor
  - Ramp response model
    - Introducing the MOSFET effective resistance
- Inverter capacitances
  - Model for scaled inverter: effective resistance and capactiance
- Inverter pair delay
  - Normalizing the inverter delay wrt τ=0.7RC
- Delay w. more than one inverter
  - Inverter pair delay
  - The fanout-of-four (FO4) delay
  - The tapered buffer, finding sizes and number of inverters
  - (Buffers/drivers with branching)

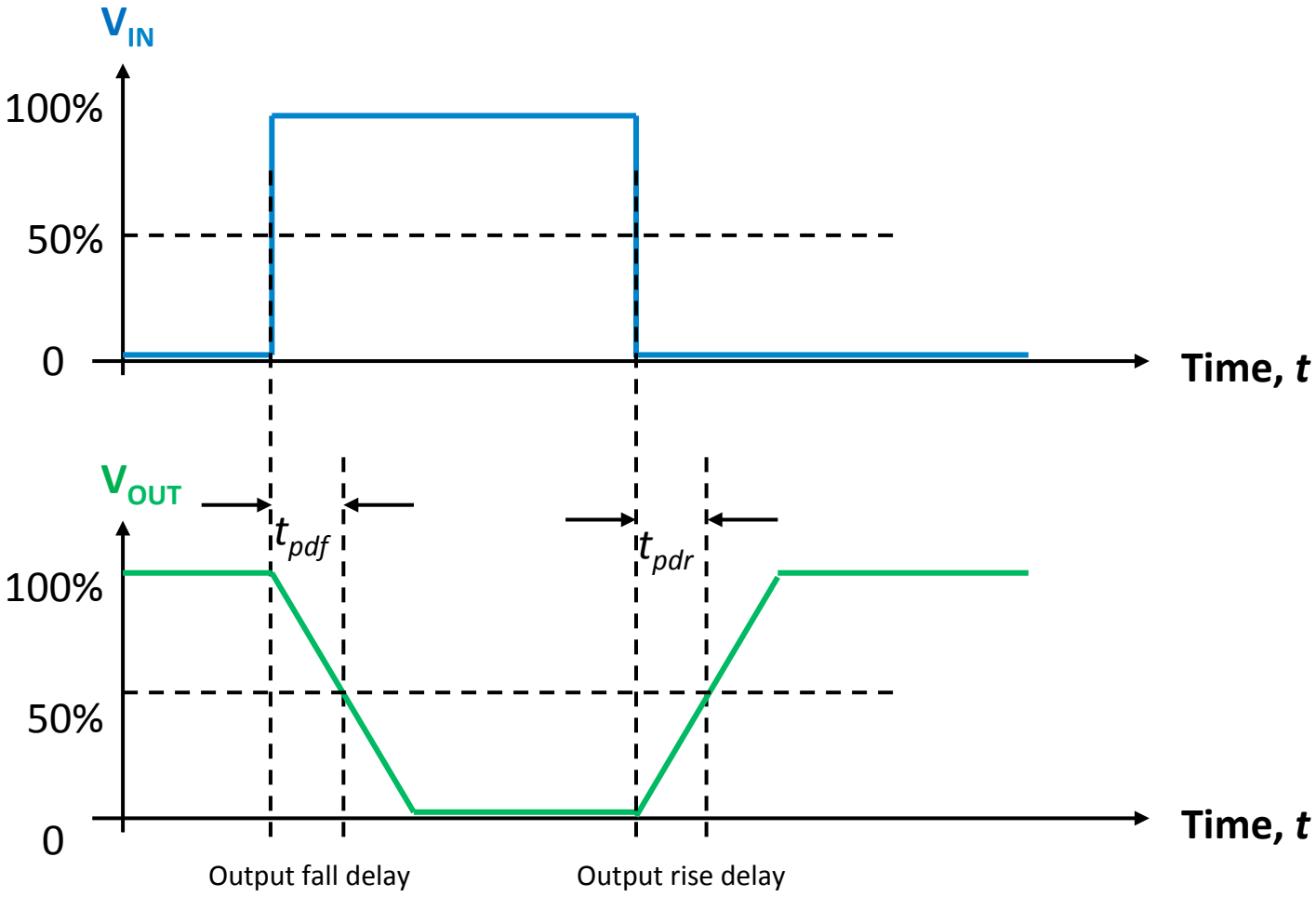
Definitions for waveforms



Definitions for delay

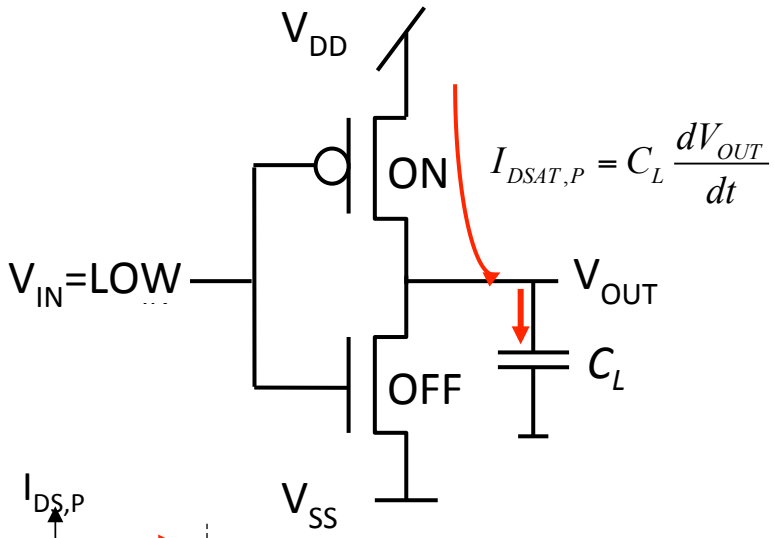


Step-response model

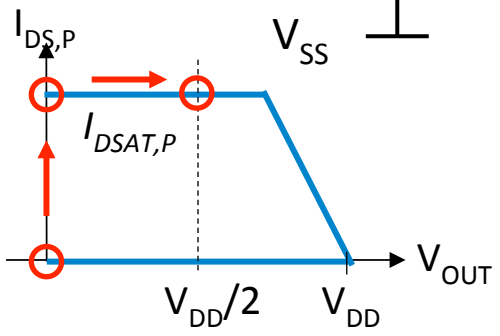


Step-response model

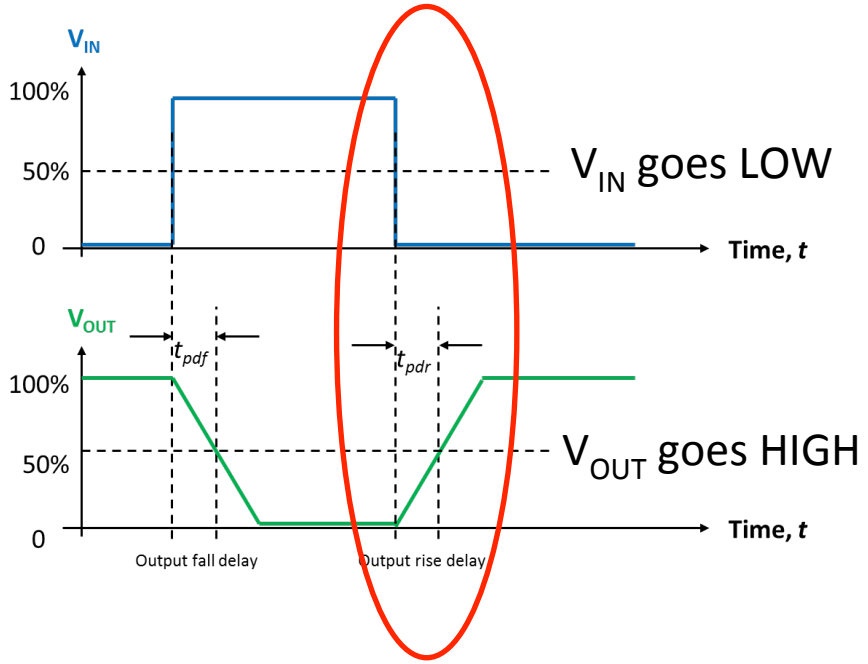
1. Charging the load capacitor through the p-channel MOSFET



$$I_{DSAT,P} = C_L \frac{dV_{OUT}}{dt}$$



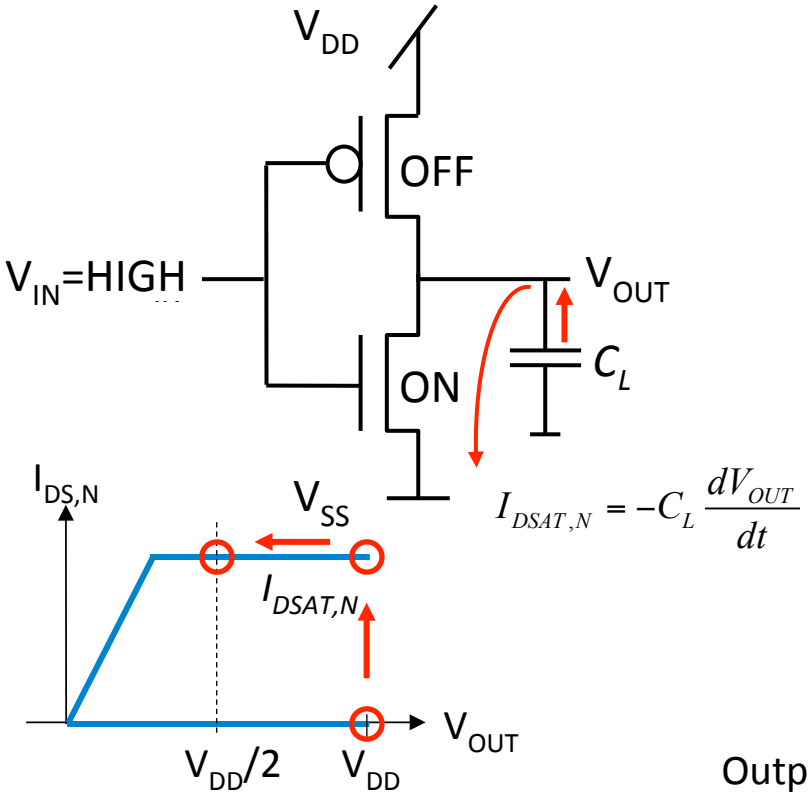
pMOS current flow in detail



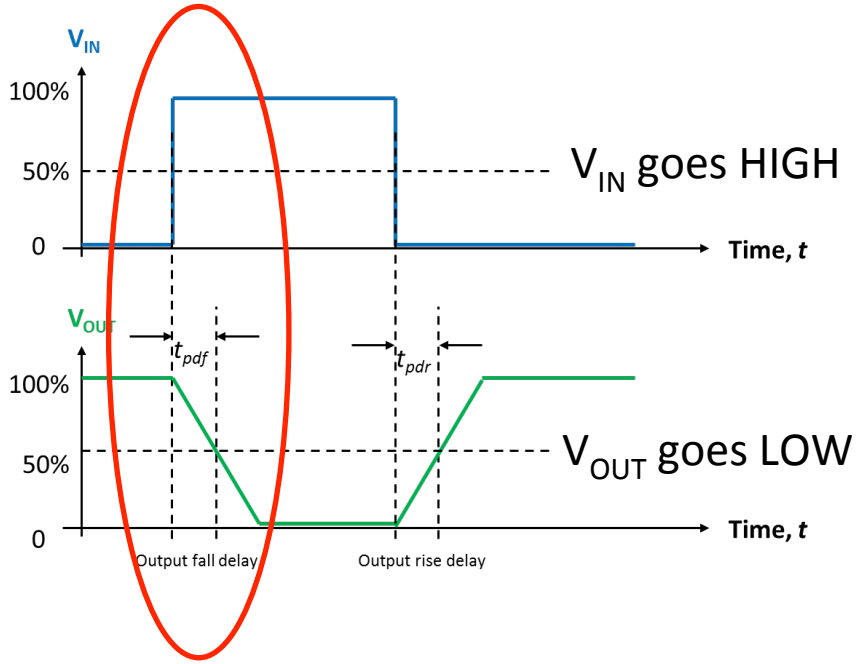
Output rise delay $t_{pdr} = C_L \frac{V_{DD}/2}{I_{DSAT,P}}$

Step-response model

2. Discharging the load capacitor through the n-channel MOSFET



$$I_{DSAT,N} = -C_L \frac{dV_{OUT}}{dt}$$

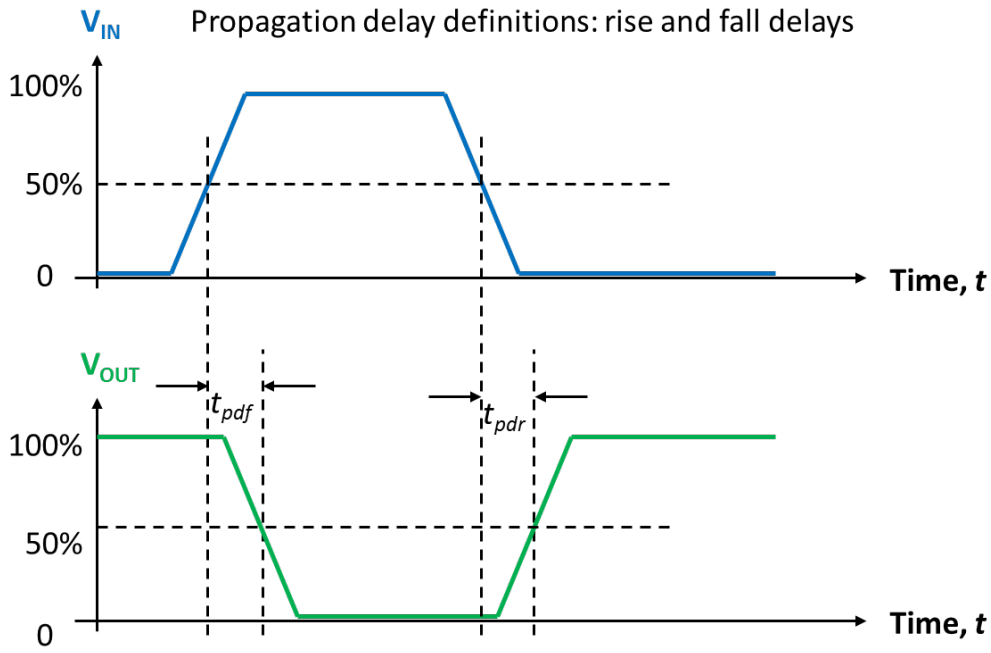


Output fall delay $t_{pdf} = C_L \frac{V_{DD}/2}{I_{DSAT,N}}$

nMOS current flow in detail

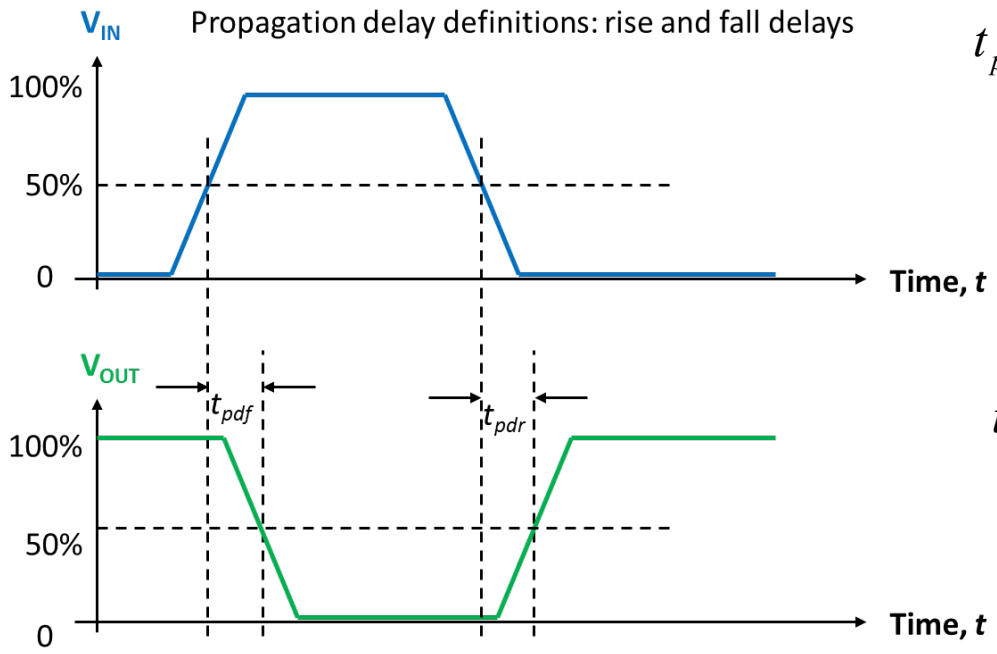
Step-response model accuracy

- How good is the step-response model?
  - Real world input voltages are not step functions
  - They are output voltages from other gates



Step response model accuracy

Experience and hundreds of circuit simulations show that propagation delays are about 40% longer in designs where input and output edge rates are equal



$$t_{pd} = C_L \frac{V_{DD}/2}{I_{DSAT}}$$

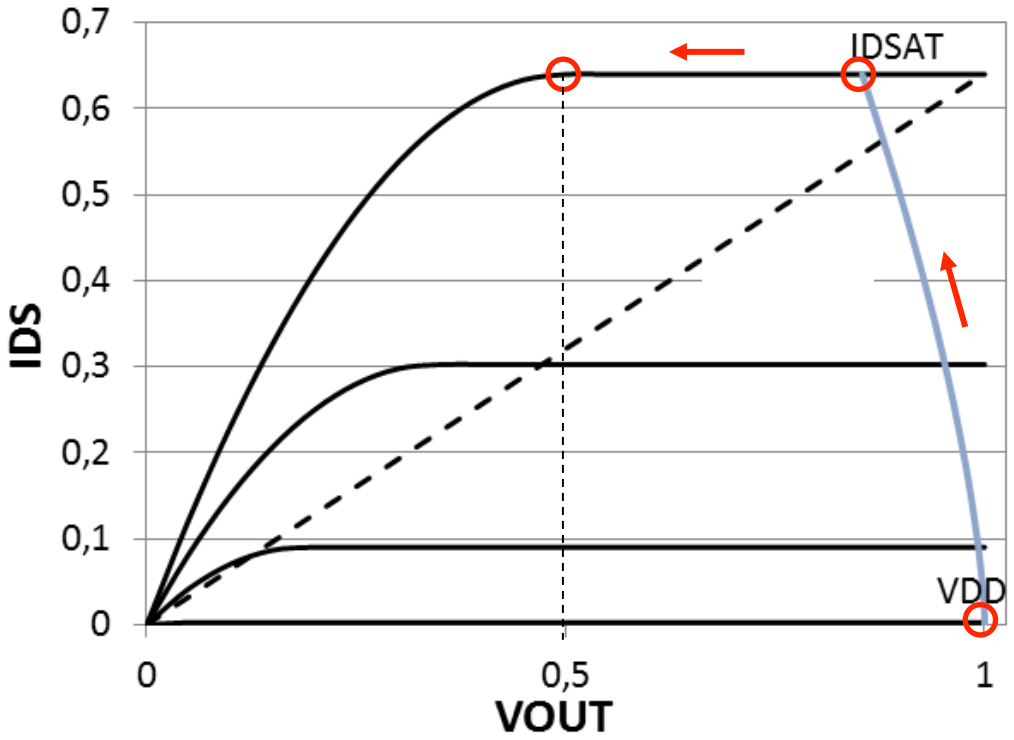
$$t_{pd} = 0.7 C_L \frac{V_{DD}}{I_{DSAT}}$$

The most important equation

$$t_{pd} = 0.7 C_L \frac{V_{DD}}{I_{DSAT}}$$

Ramp input – output trace

V_{IN}:  V_{GS} rises from 0 to V_{DD}



$V_{GS} = V_{DD}$

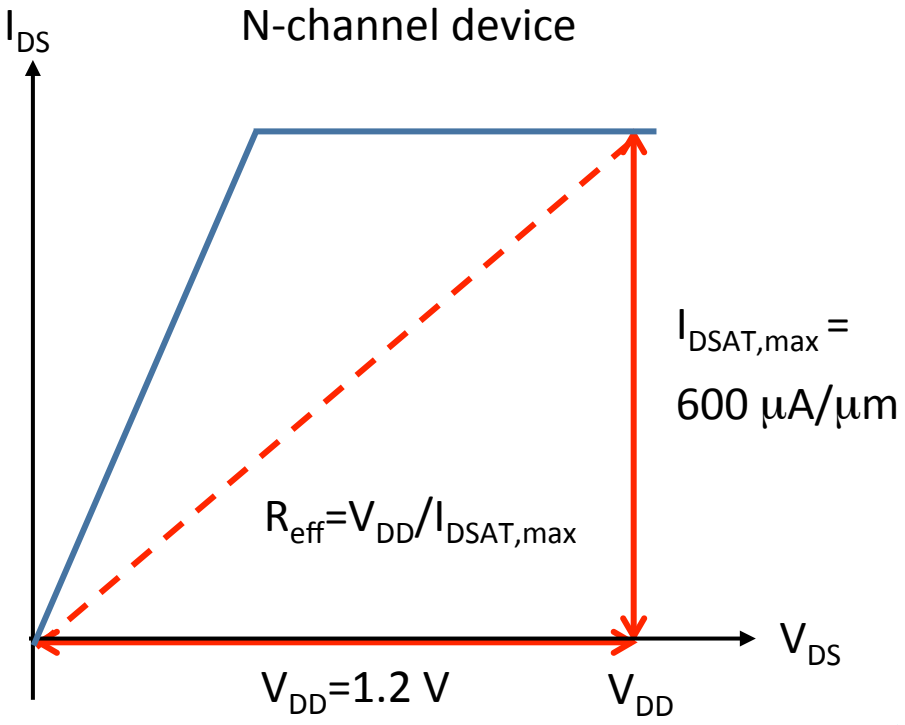
$V_{GS} = 0$

Some discharge happens while input rises from 0 to VDD
Most discharge still happens with IDSAT,max

Effective resistances: 65 nm MOSFETs

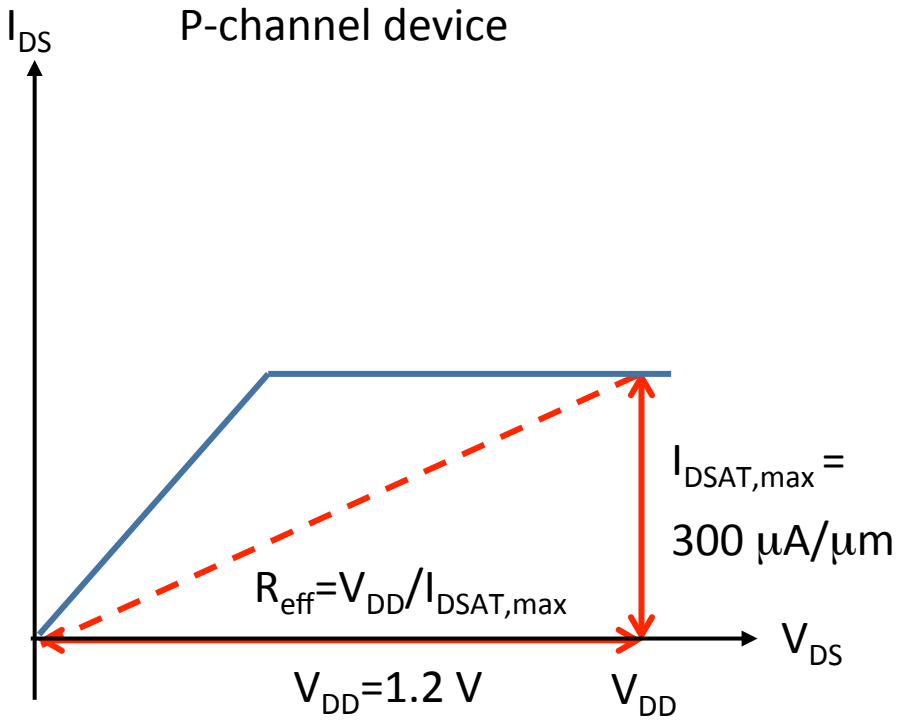
$R_{N,eff}$=2 kΩ·μm

N-channel device



$R_{P,eff}$=4 kΩ·μm

P-channel device



$$t_{pd} = 0.7 C_L \frac{V_{DD}}{I_{DSAT}} \Rightarrow t_{pd} = 0.7 R_{eff} C_L$$

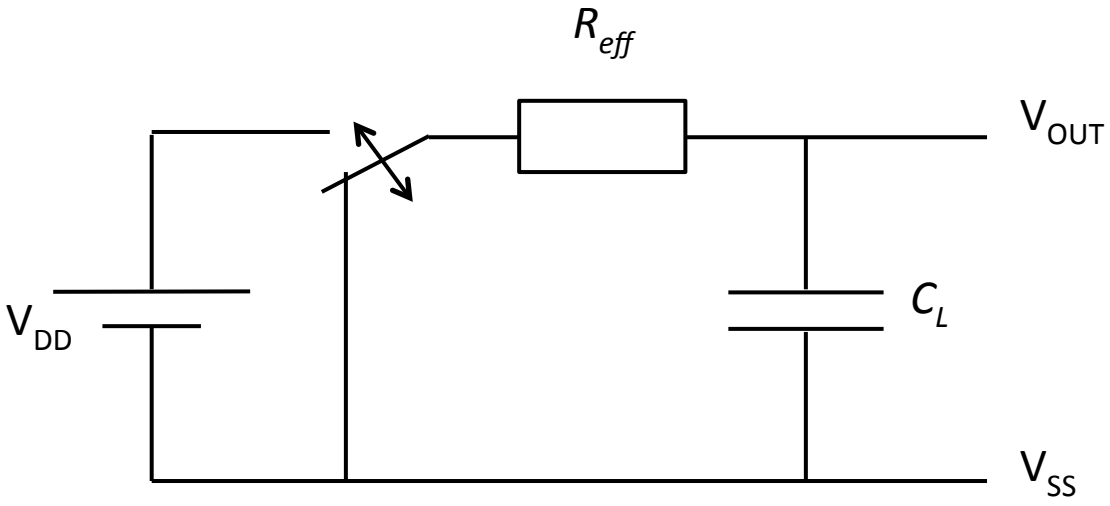
The most important equation revisited

$$t_{pd} = 0.7 C_L R_{eff}$$

RC delay

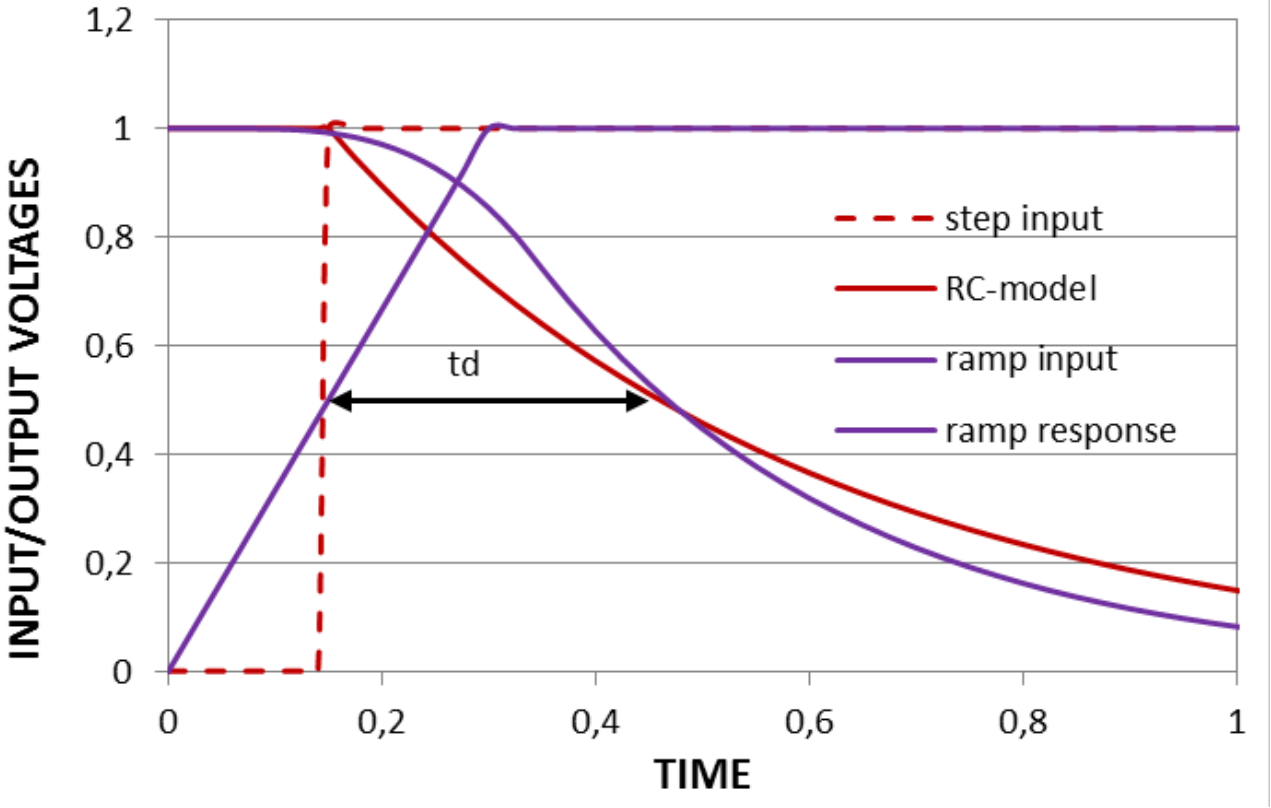
This RC circuit has an output voltage delay given by

$$t_{pd} = 0.7 R_{eff} C_L$$



Ramp input – output response

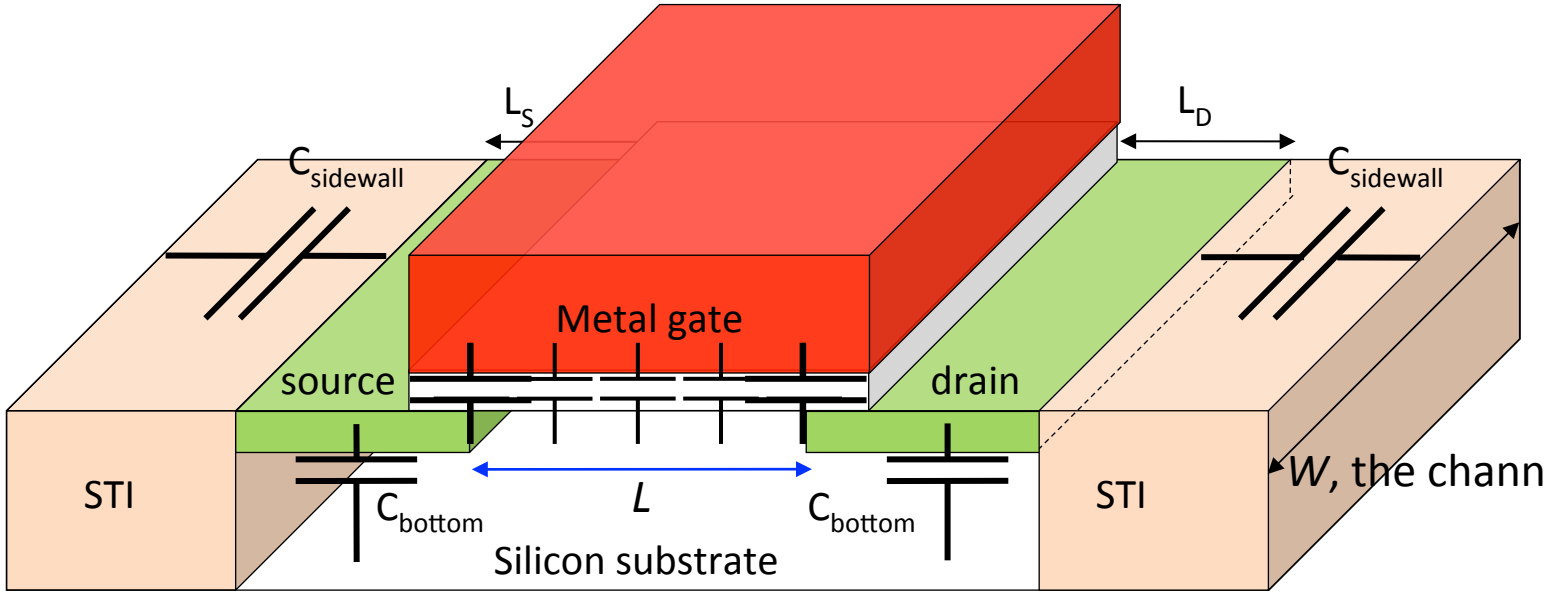
The two curve forms are not the same, but they yield the same delay!



MOSFET capacitances (from lecture 2)

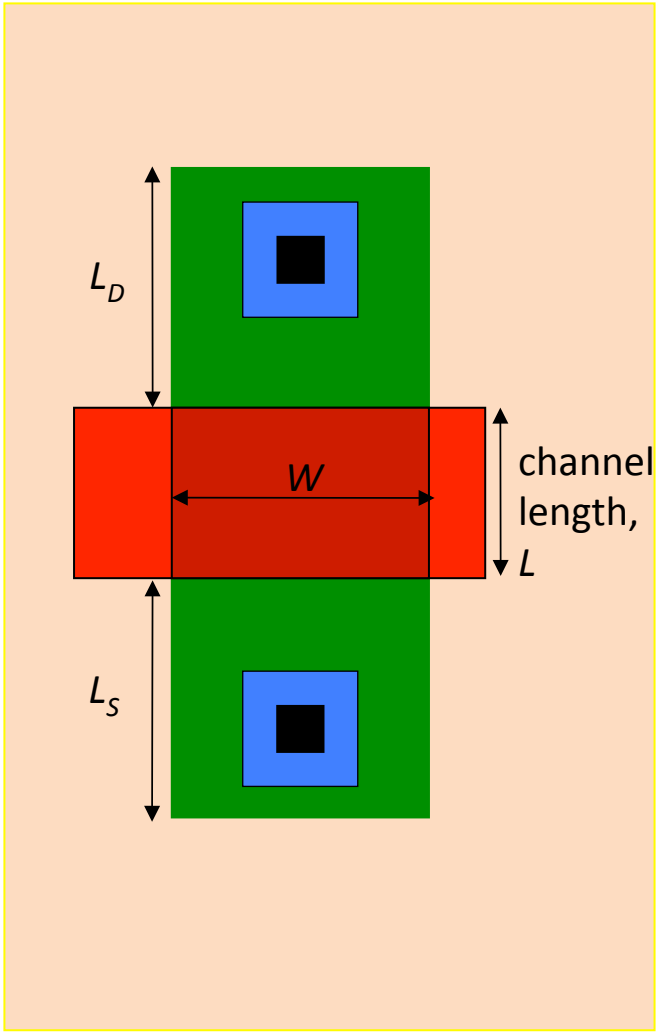
There are parasitic capacitances from source/drain to substrate

There is a distributed gate capacitance that must be lumped to to the available ckt nodes, that is, source and drain



L, the channel length

65 nm CMOS capacitances



Gate capacitance

Plate capacitance between gate and channel (but here we model it between gate and source)

$C_G = WLC_{ox}$

$C_{ox} \approx 20$ fF/µm^2

(corresponds to 1.2 nm effective oxide thickness, t_{ox})

60 nm effective channel length yields:

$C_G = 1.2$ fF/µm

Drain/source parasitic capacitances

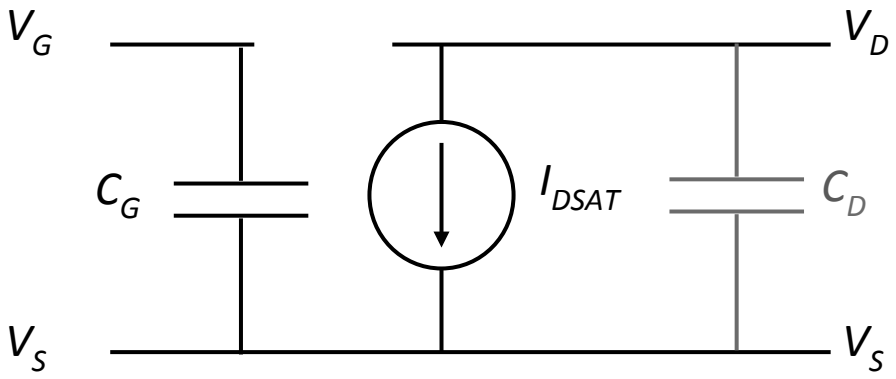
Same order of magnitude as the gate capacitance,

$C_S = C_D = p_{inv} C_G$

Generic approach is to use $p_{inv}=1$ for simplicity

But sometimes $p_{inv}=0.5$ or $p_{inv}=0.8$!

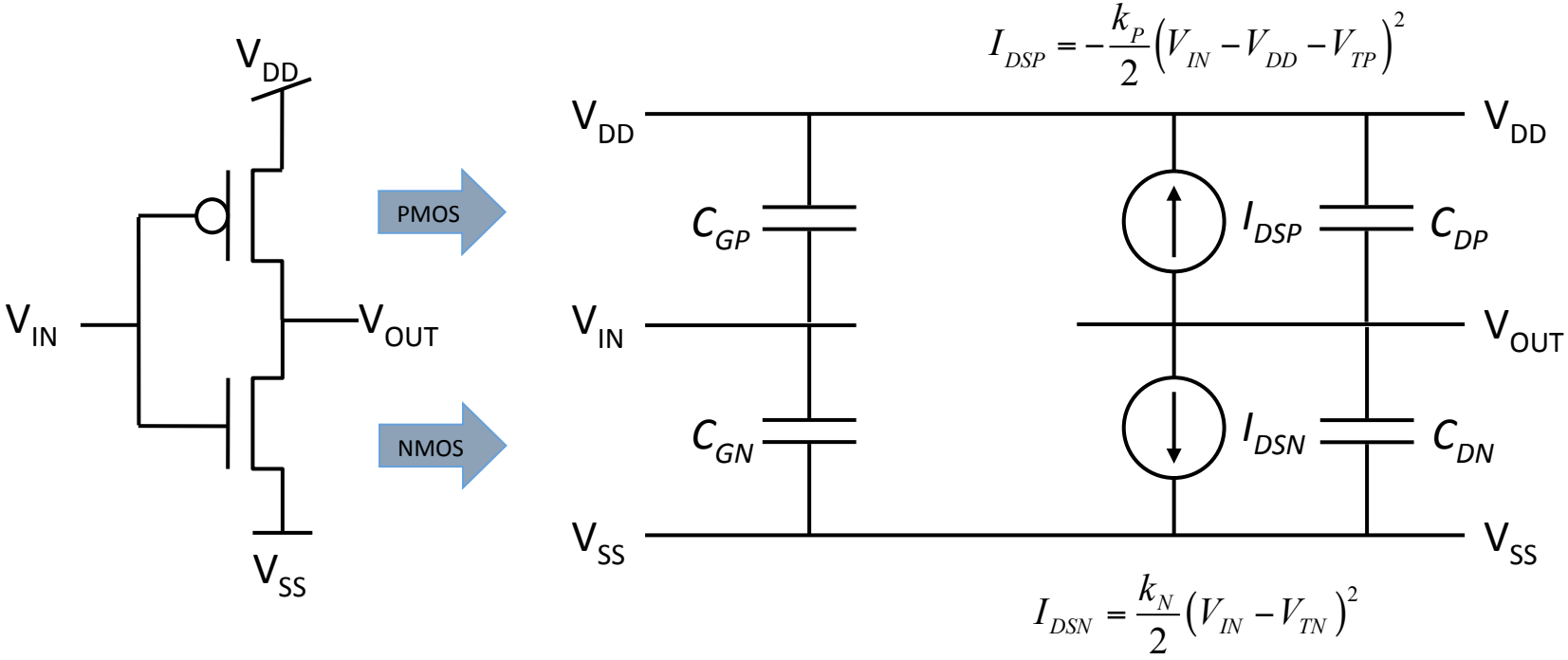
Electrical model for saturated MOSFET (from lecture 2)



Model for saturated MOSFET
Valid for both nMOS and pMOS transistors

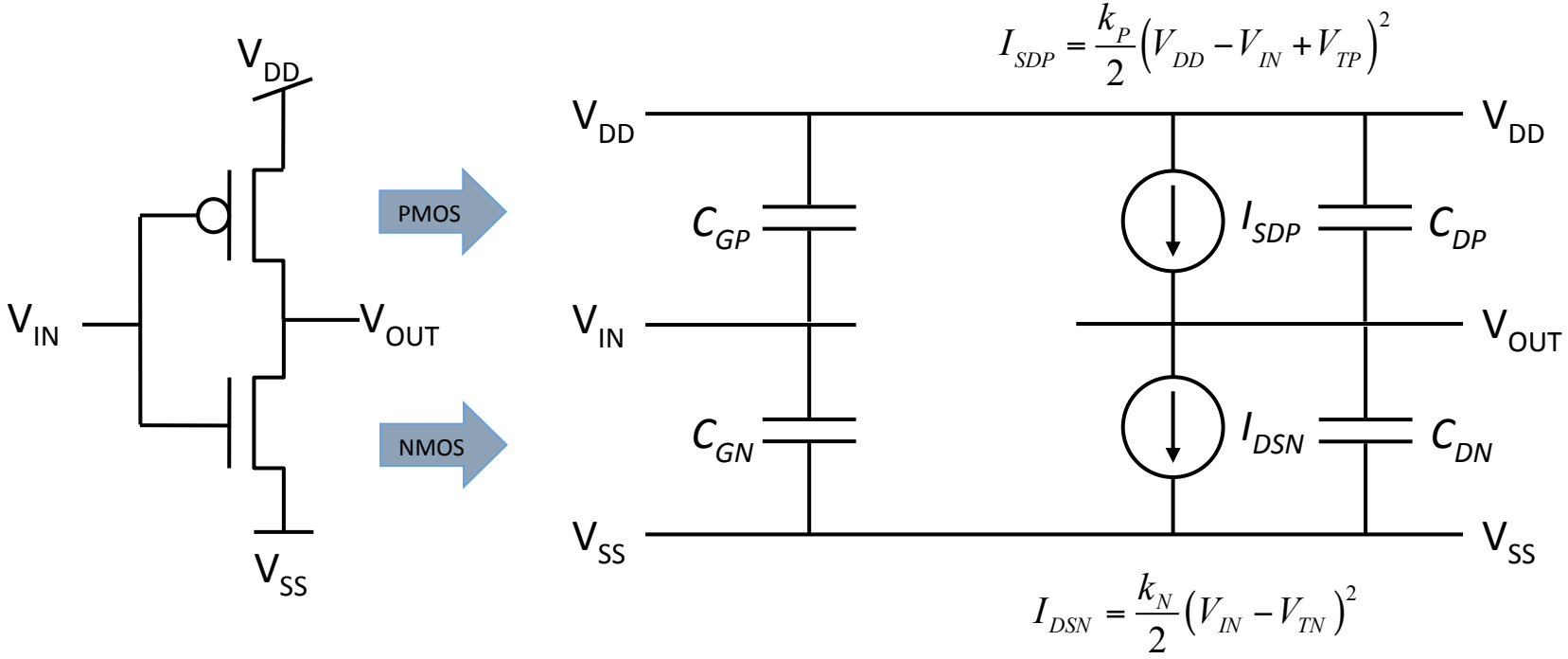
The inverter and its electrical model

Replace the MOSFETs with their equivalent electrical circuits!



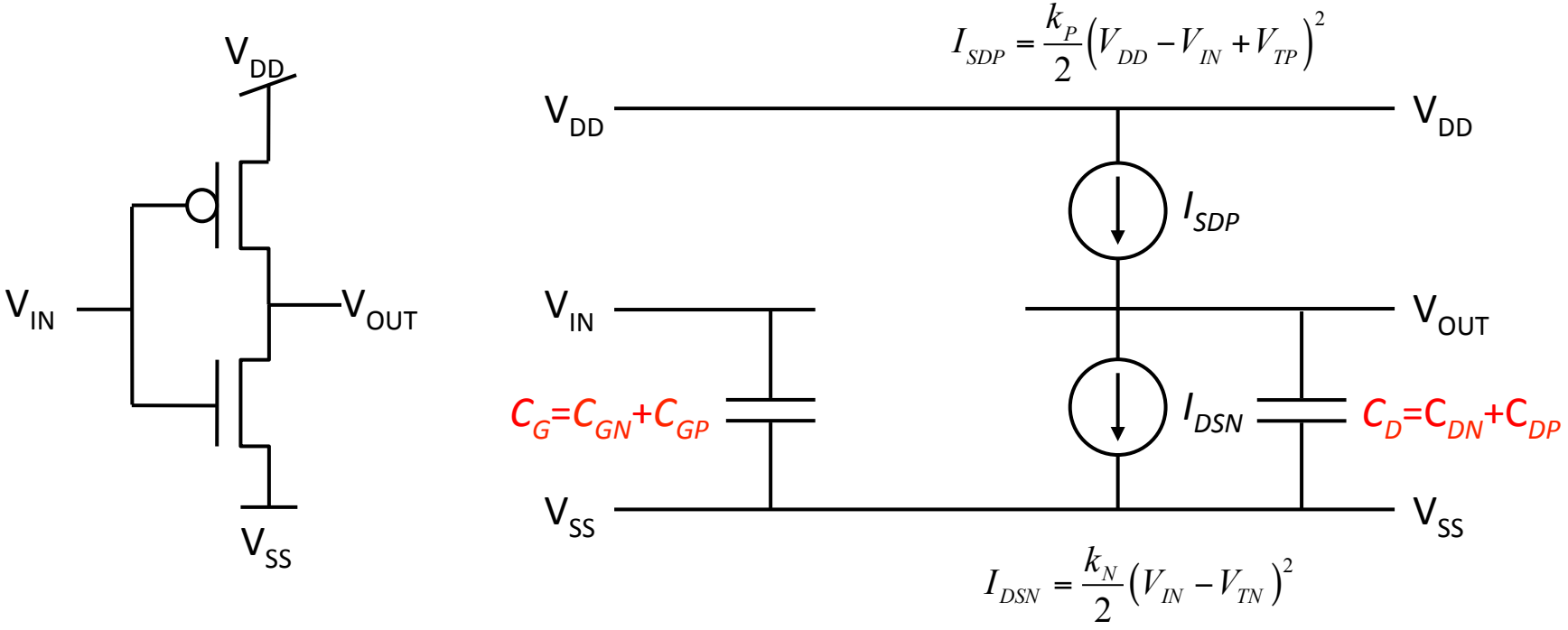
The inverter and its electrical model

Change the sign of the pMOS current so that all currents are positive
(We also made the pMOS voltages positive, although it is not necessary)



The inverter and its electrical model

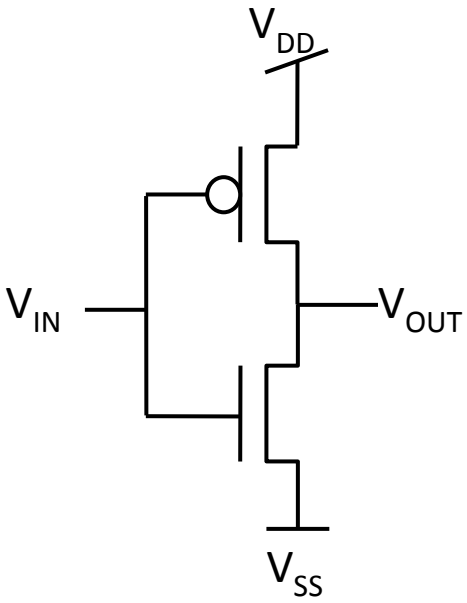
Place all capacitors to signal ground! Both rails are constant voltages, no dV/dt

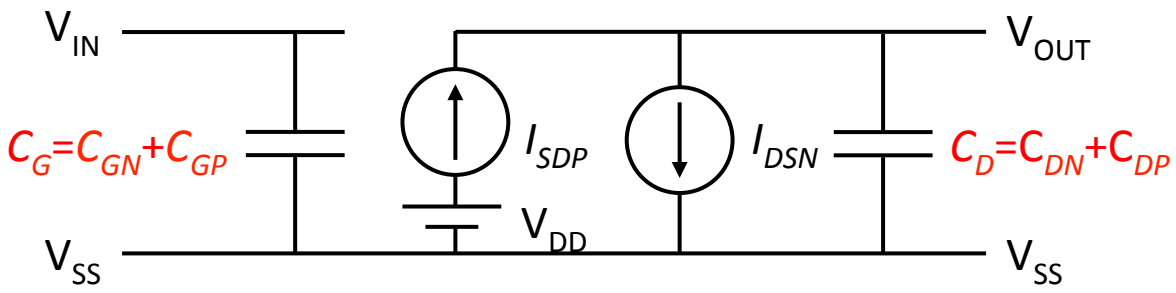


Inverter input capacitance: $C_G = C_{GN} + C_{GP}$; MOSFET gate capacitances add!
Inverter parasitic output capacitance: $C_D = C_{DN} + C_{DP}$. Drain caps also add!

The inverter and its electrical model

Eliminate VDD rail by inserting power supply to signal ground!

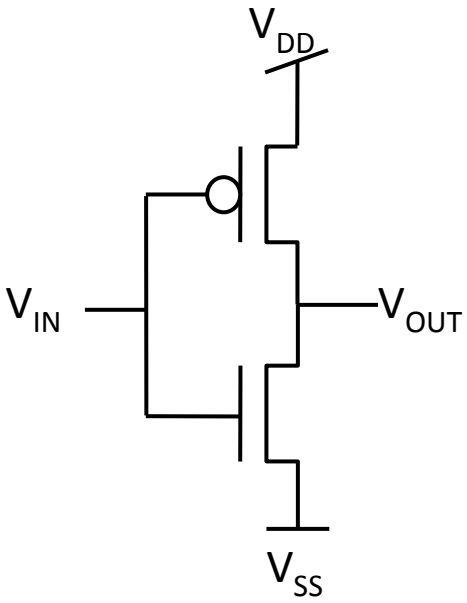




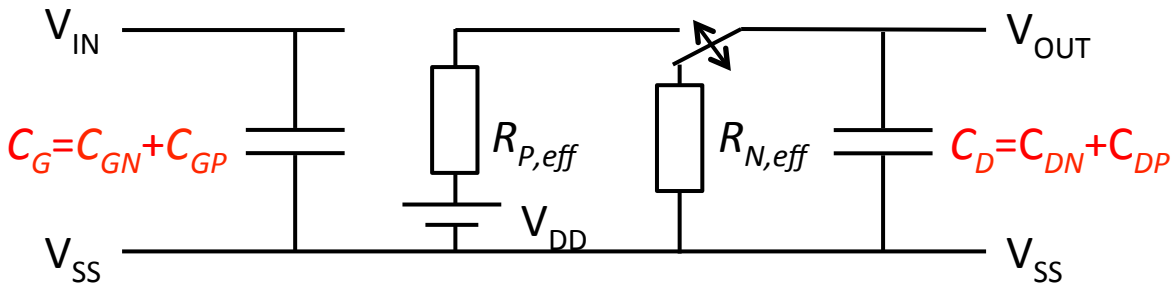
Inverter input capacitance: $C_G=C_{GN}+C_{GP}$; MOSFET gate capacitances add!
Inverter parasitic output capacitance: $C_D=C_{DN}+C_{DP}$. Drain caps also add!

The inverter and its electrical model

Replace MOSFET constant-current sources with their effective resistances!



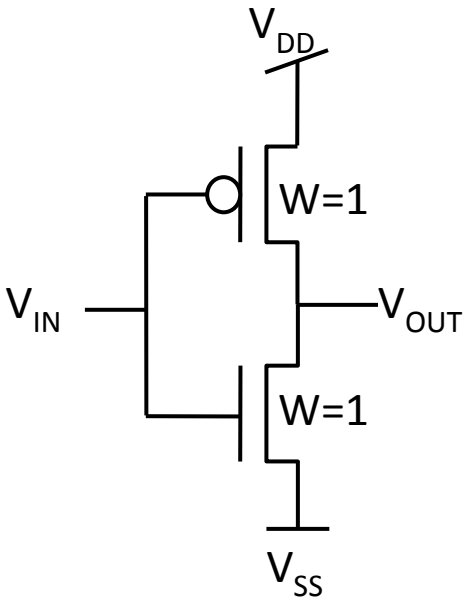
$R_{N,eff}$=2 kΩ·μm $R_{P,eff}$=4 kΩ·μm



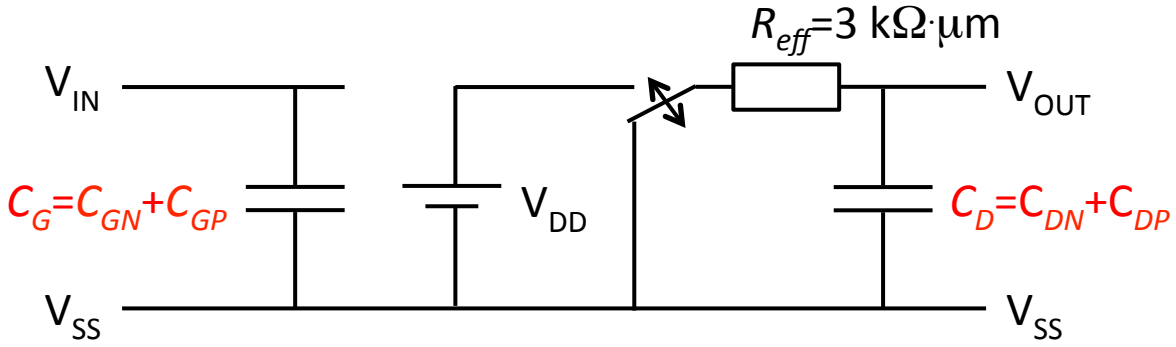
Inverter input capacitance: $C_G=C_{GN}+C_{GP}$; MOSFET gate capacitances add!
Inverter parasitic output capacitance: $C_D=C_{DN}+C_{DP}$. Drain caps also add!

The inverter and its electrical model

Too cumbersome to have different rise and fall delays!
Replace effective resistances with one average effective resistance!



$R_{N,eff}$=2 kΩ·μm $R_{P,eff}$=4 kΩ·μm

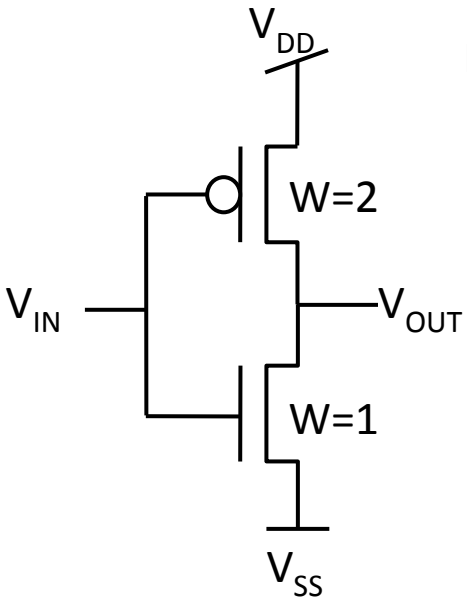


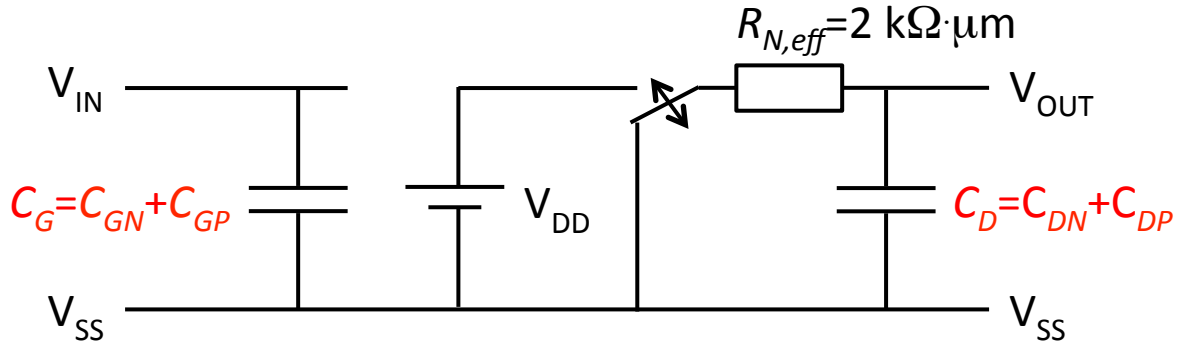
Inverter input capacitance: $C_G=C_{GN}+C_{GP}$; MOSFET gate capacitances add!
Inverter parasitic output capacitance: $C_D=C_{DN}+C_{DP}$. Drain caps also add!

The inverter and its electrical model

Too cumbersome to have different rise and fall delays!
Replace effective resistances with one average effective resistance!

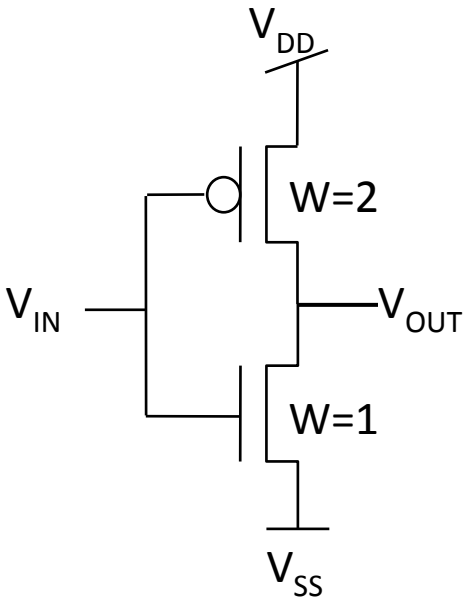
Or even better!
Design the inverter for equal effective resistances, $R_{P,eff}=R_{N,eff}$,
by making p-channel MOSFET twice as wide as the n-channel
MOSFET to compensate for the lower hole mobility



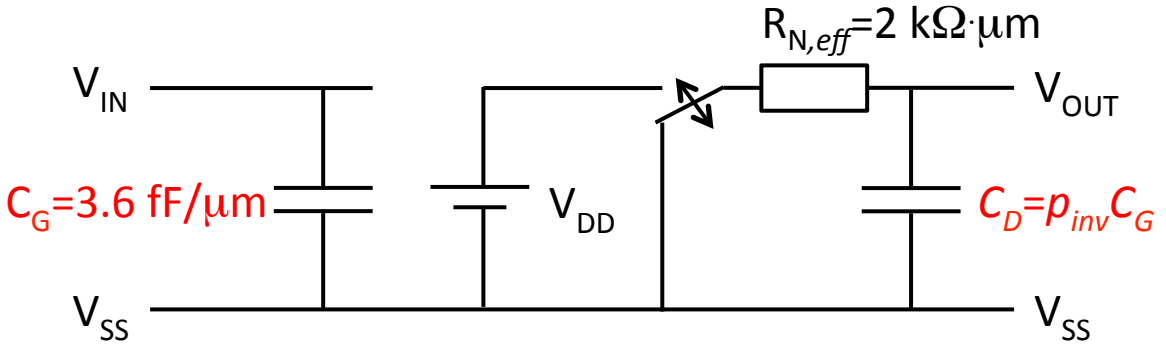


Inverter input capacitance: $C_G=C_{GN}+C_{GP}$; MOSFET gate capacitances add!
Inverter parasitic output capacitance: $C_D=C_{DN}+C_{DP}$. Drain caps also add!

Inverter capacitances



Task: Calculate C_G and C_D with pMOS FET is twice as wide as nMOS FET expressed in the width of the nMOS transistor.



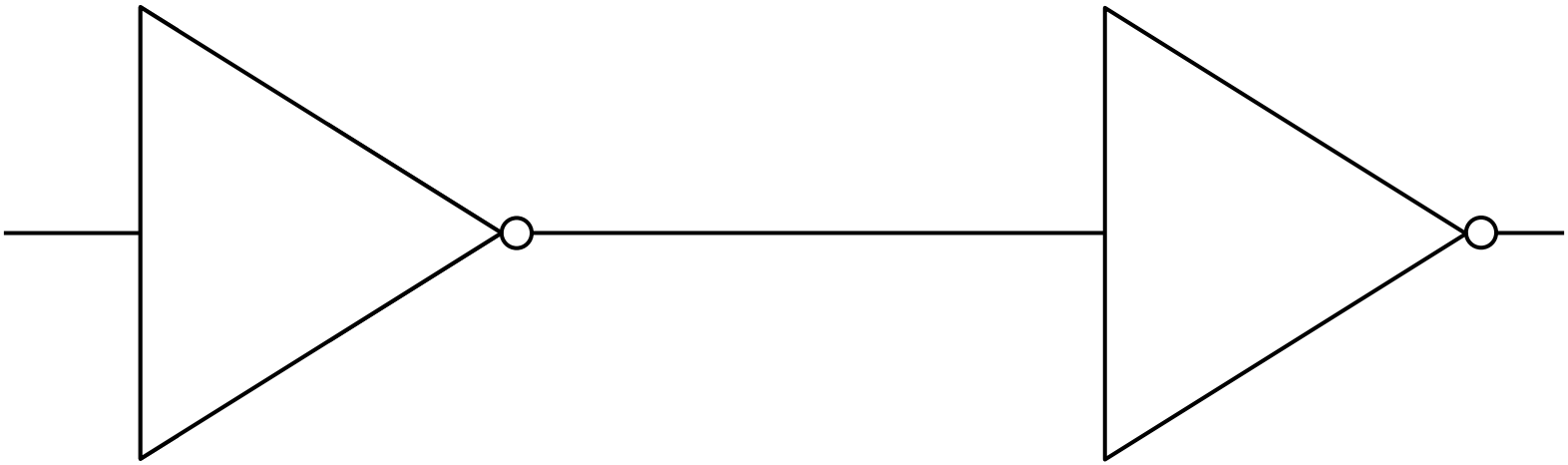
We notice that $C_G \times R_{N,eff}$ is a constant. It will soon return!

Answer: Assuming L=60 nm and C_{ox}=20 fF/μm² we obtain C_{GN}=1.2 fF/μm and C_{GP}=2.4 fF/μm. Hence C_G=3.6 fF/μm.
Concerning C_D we assume $C_D=p_{INV}C_G$=3.6 fF/μm with p_{INV}=1.

More than one inverter

Inverter pair delay

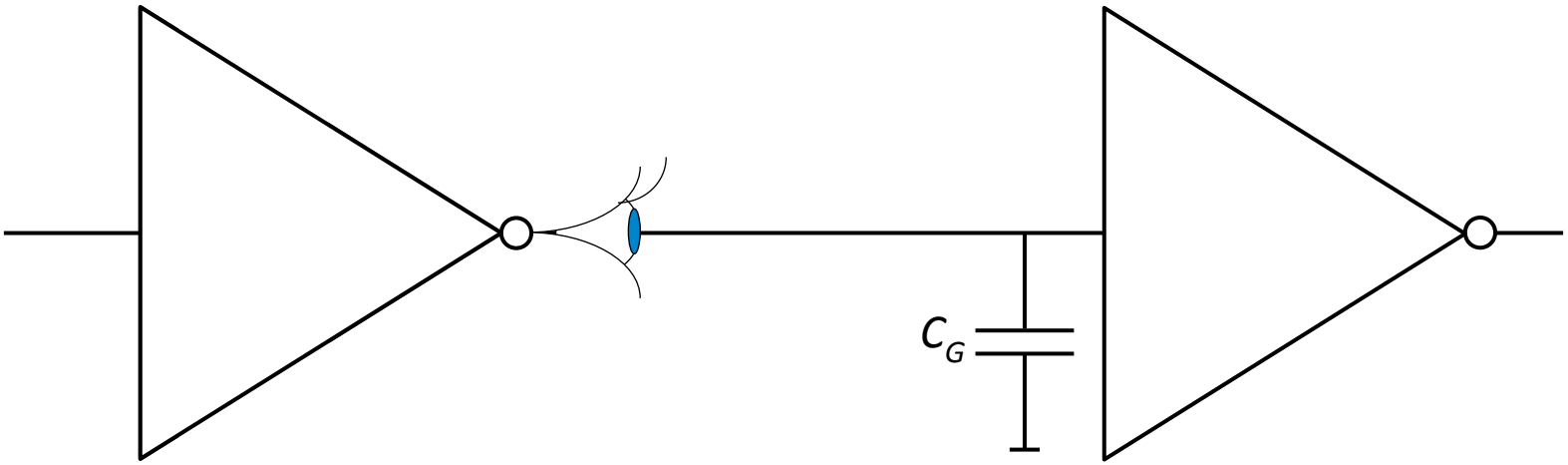
Task: Calculate the inverter pair delay!



Inverter pair delay

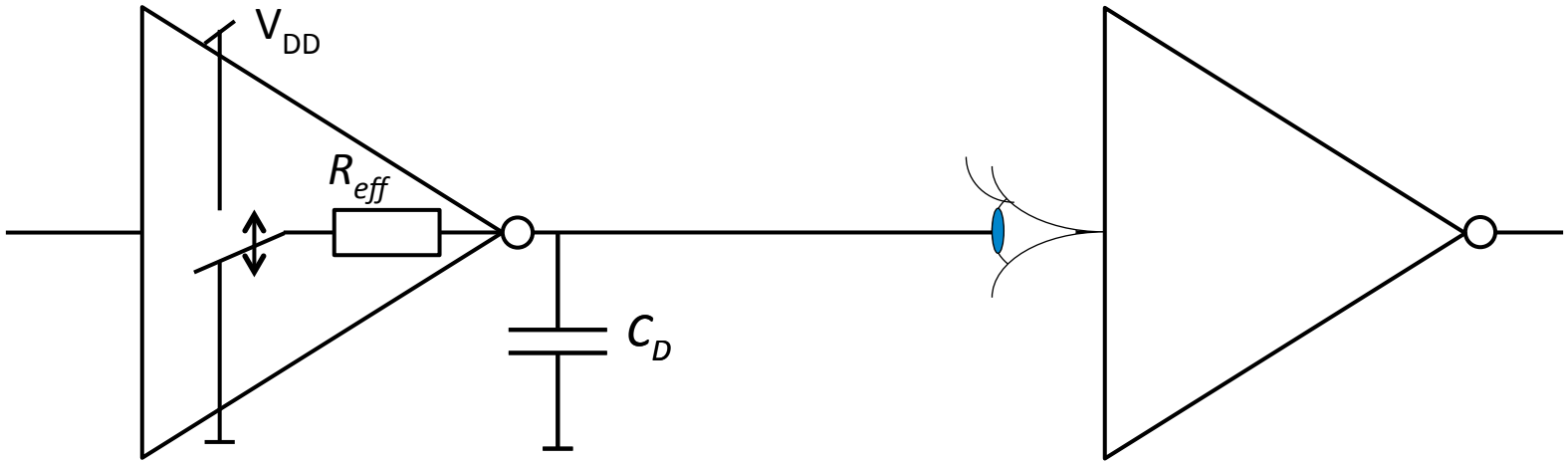
Put on your “two-port glasses” and look towards the loading inverter!

You will only see the input capacitance of the loading inverter



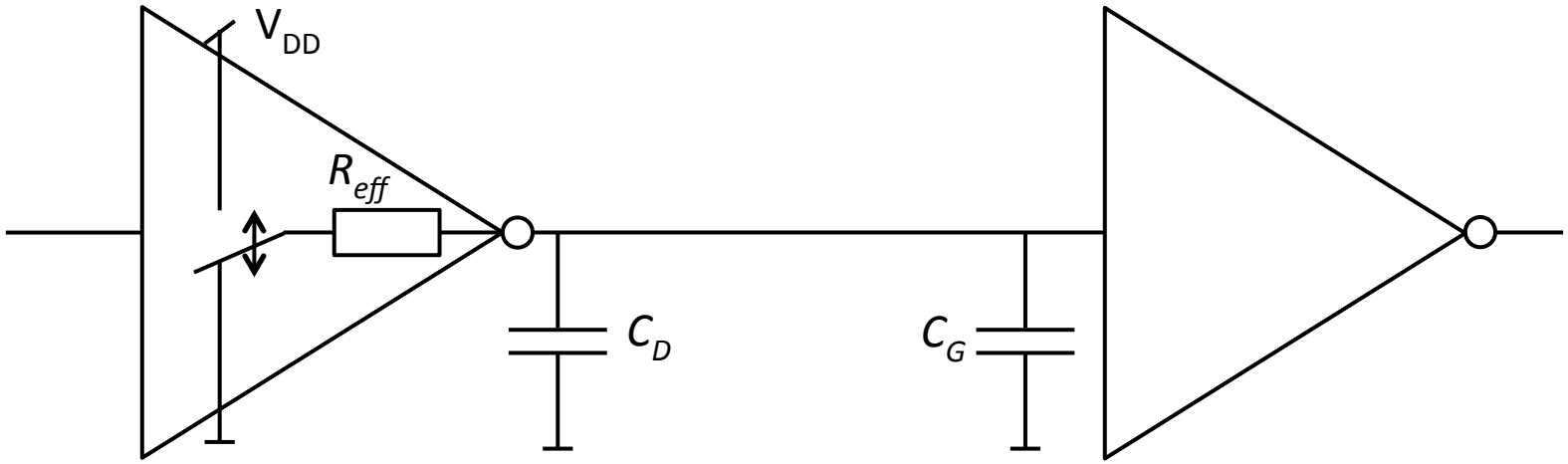
Inverter pair delay

Have your “two-port glasses” on and look towards the driving inverter!
You will see a voltage source with a certain source resistance, and you will see the parasitic capacitance of the loading inverter



Inverter pair delay

In an ideal inverter the time constant tau is really what the name says, a constant;
It is independent of inverter size (as long as $W_P/W_N=2$).



Propagation delay:

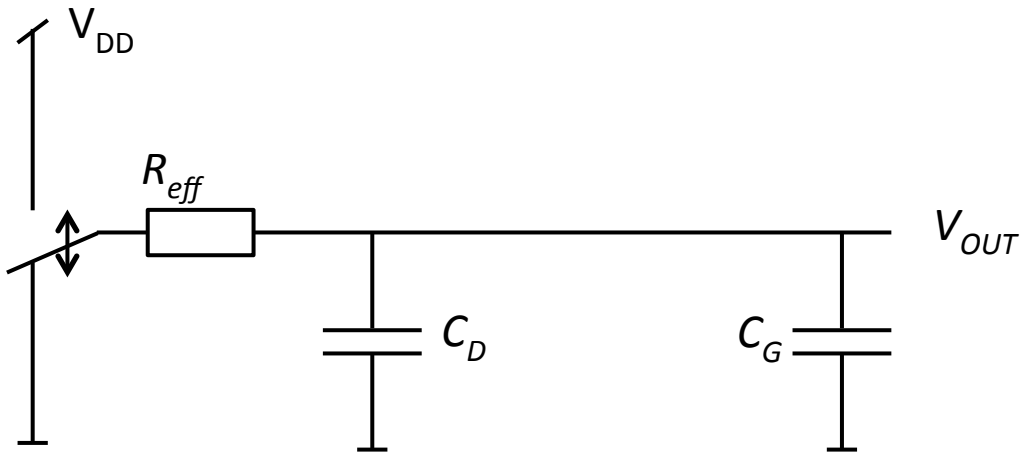
$$t_{pd} = 0.7R_{eff}\left(C_D + C_G\right) = 0.7R_{eff}C_G\left(p_{inv} + 1\right) = 5\ \text{ps} \times 2 = 10\ \text{ps}$$

All delay calculations are made wrt to this technology time constant tau (τ)

$$\tau = 0.7R_{eff}C_G = 0.7 \times \left(2\ \text{k}\Omega \times \mu\text{m}\right) \times \left(3.6\,fF/\mu\text{m}\right) = 5\ \text{ps}$$

Inverter pair delay

Equivalent electrical circuit for propagation delay calculations



Propagation delay:

$$t_{pd} = 0.7R_{eff}\left(C_D + C_G\right) = 0.7R_{eff}C_G\left(p_{inv} + 1\right) = 5\ \text{ps} \times 2 = 10\ \text{ps}$$

All delay calculations are made wrt to this technology time constant tau (τ)

$$\tau = 0.7R_{eff}C_G = 0.7 \times \left(2\ \text{k}\Omega \times \mu\text{m}\right) \times \left(3.6\,fF/\mu\text{m}\right) = 5\ \text{ps}$$

Important concept

The technology time constant tau, τ:

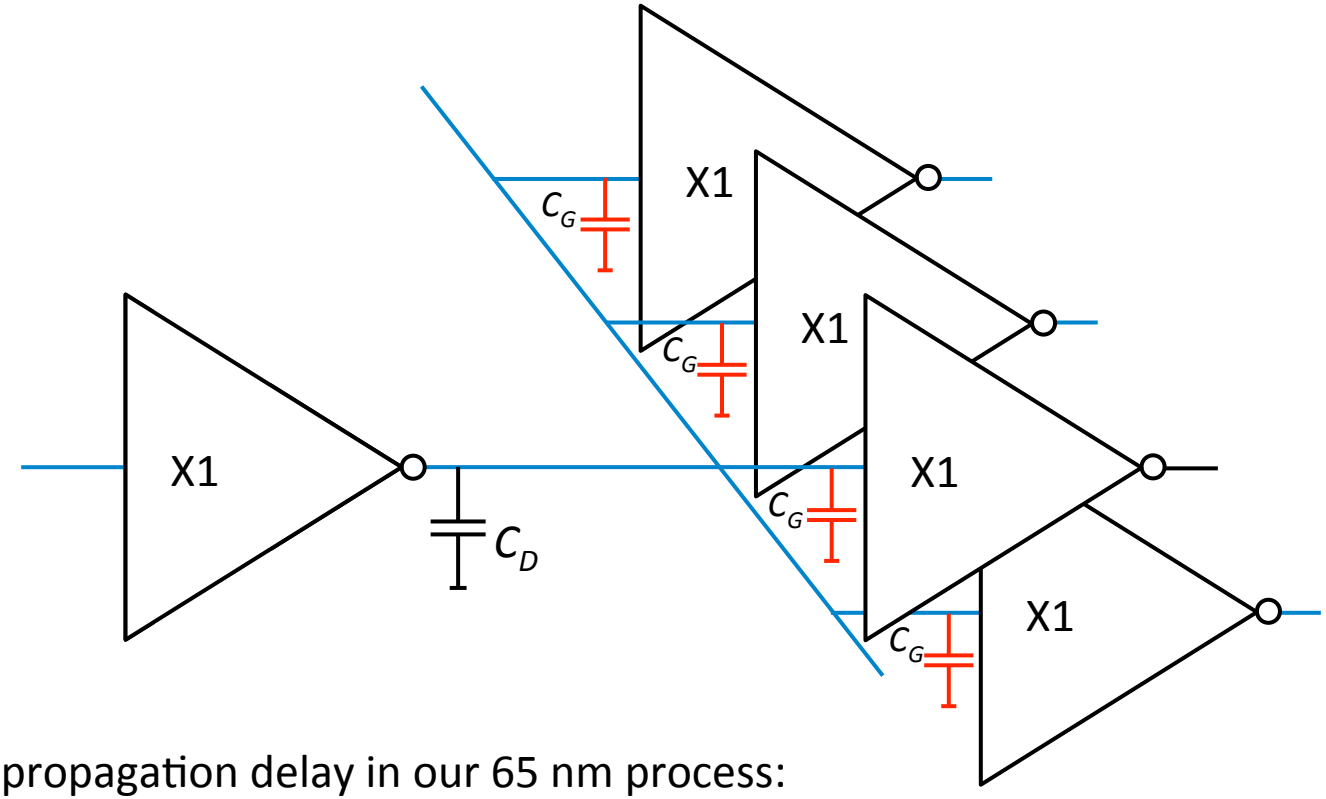
$$\tau = 0.7 R_{\text{eff}} C_G$$

Quiz time!

- Go to socrative.com
- Select Student login
- Go to room: “MCC0922018”

Inverter FO4 delay

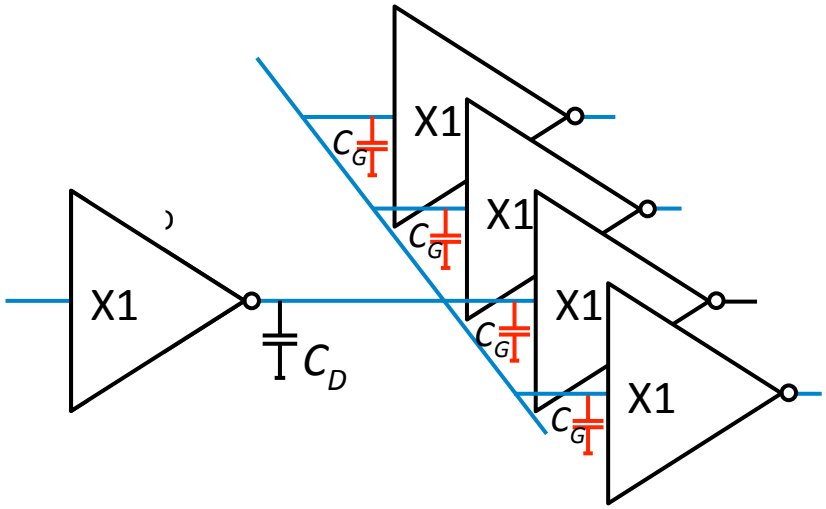
One inverter drives 4 identical inverters



The FO4 propagation delay in our 65 nm process:

$$t_{pd} = 0.7R_{eff}\left(C_D + 4C_G\right) = \underbrace{0.7R_{eff}C_G}_{\text{tau}}\left(p + 4\right) = 5\text{ ps} \times 5 = 25\text{ ps}$$

Normalized FO4 delay



$$\tau = 0.7 R_{\text{eff}} C_G$$

$$t_{pd} = (p_{\text{inv}} + 4)\tau \approx 5\tau$$

Normalized delay means normalized to τ

$$d \equiv \frac{t_{pd}}{\tau}$$

So FO4 normalized delay: $d \approx 5$

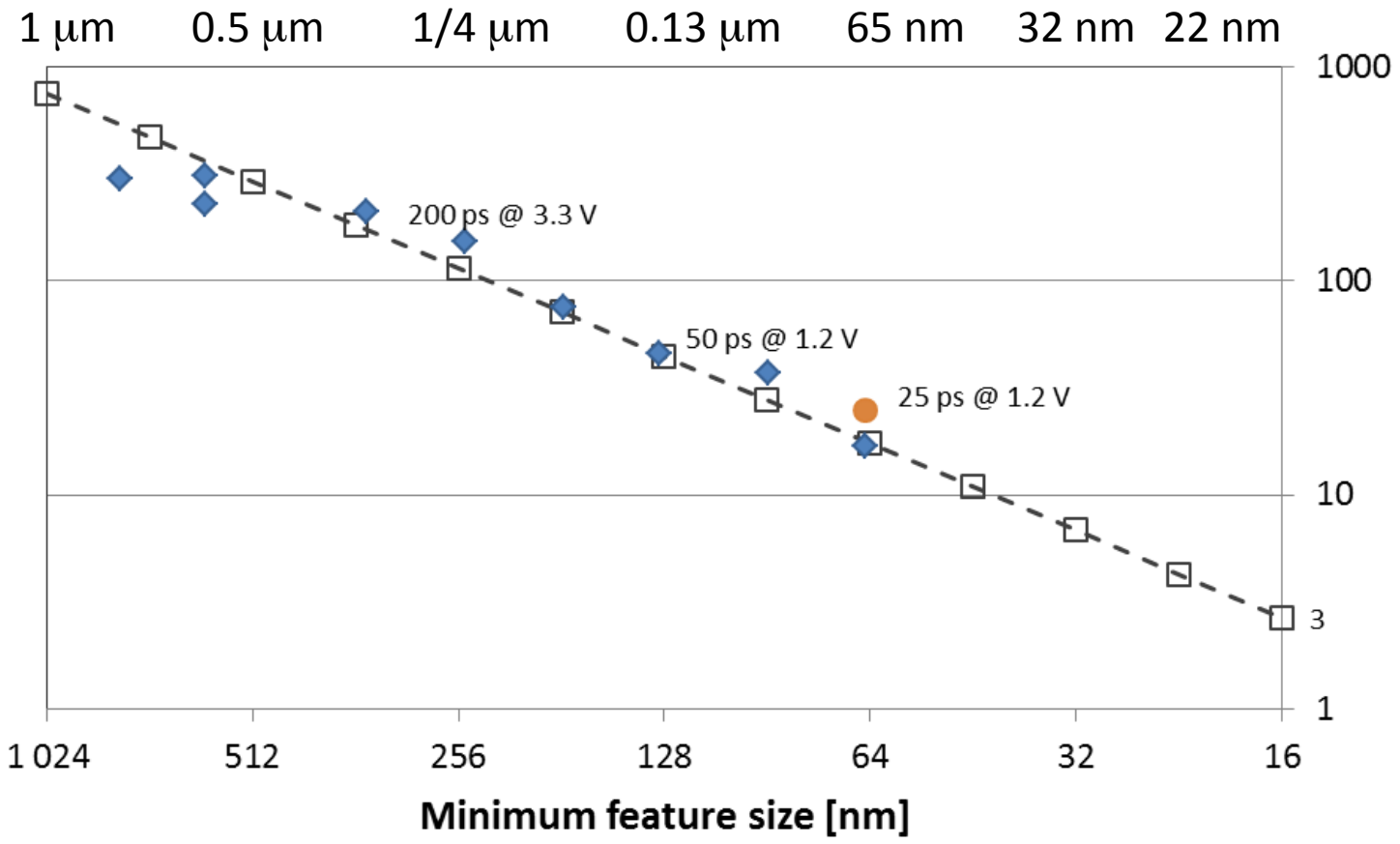
Important concept

The normalized delay, d:

$$d \equiv \frac{t_{pd}}{\tau}$$

Where tau, τ, is the technology time constant, τ

FO4 delay trends vs. feature size



Microprocessor cycle time trends

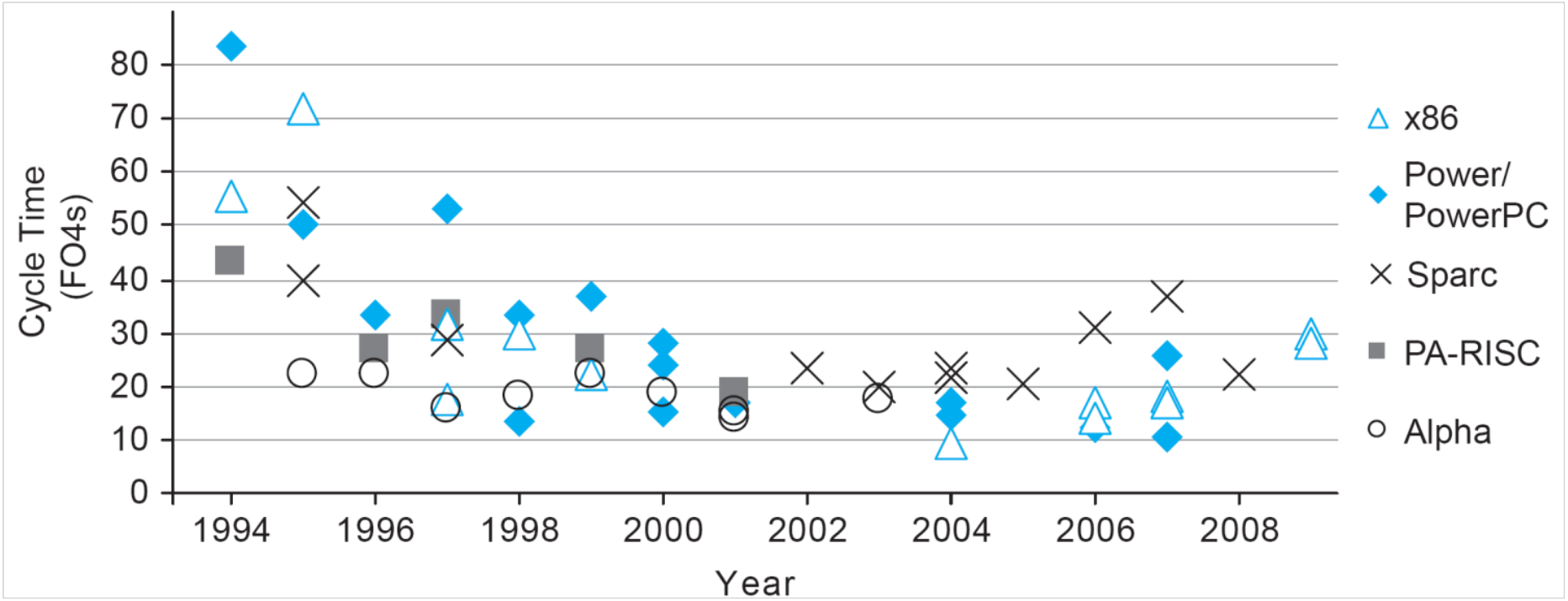


FIGURE 4.38 Microprocessor cylcle time trends. Data has some uncertainty based on estimating FO4 delay as a function of feature size.

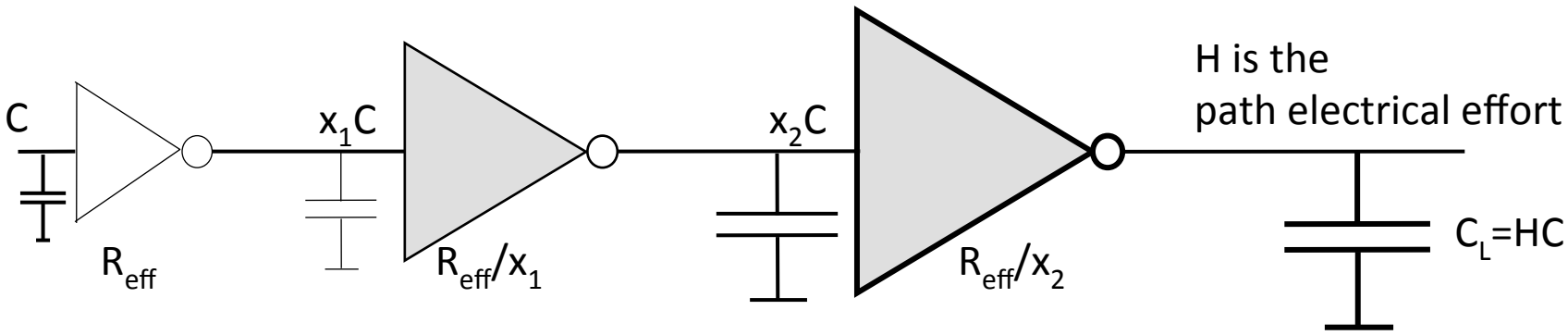
From Weste & Harris.

Inverter Size

- In most vendor cell libraries, inverters and other logic gates come in a number of different varieties concerning their driving capability (R_{eff}) and input capacitance (C_G).
- In the following all inverters and logic gates of a certain size, e.g. size X=8, will have the same input capacitance, and, as an example, this input capacitance will be only half the input capacitance of a gate of size X=16.

Minimizing delay through multiple inverters

Path definitions



Path electrical effort: $H = C_L/C_{IN}$

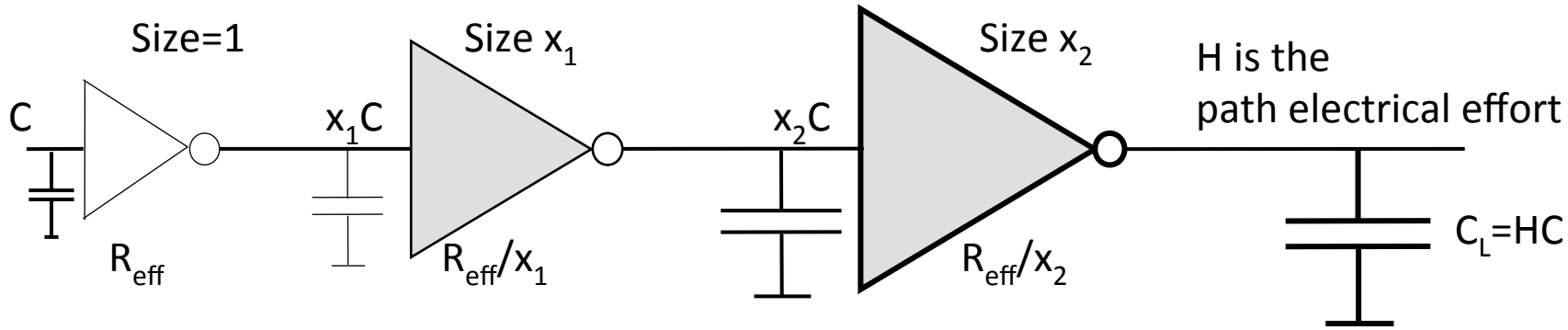
Stage electrical effort or **fanout** for inverter n: $h_n = C_{INn+1}/C_{INn} = x_{n+1}/x_n$

Normalized delay for inverter n: $d_n = h_n + p_{inv}$

Normalized path delay with N stages : $D = N \times p_{inv} + h_1 + h_2 + \ldots + h_N$

The tapered buffer with 3 stages

Reference inverter . . . and two inserted buffer inverters



Normalized path delay: $D = 3\,p_{inv} + h_1 + h_2 + h_3$ where $h_1=x_1$, $h_2=x_2/x_1$ and $h_3=H/x_2$.
But only h_1 and h_2 are independent variables, h_3 becomes $h_3=H/h_1h_2$:

$$D = 3\,p_{inv} + h_1 + h_2 + H/h_1h_2$$

Show that minimum delay is obtained for $h_1=h_2=\sqrt[3]{H}$, which gives:

$$D_{min} = 3\,p_{inv} + 3 \times \sqrt[3]{H}$$

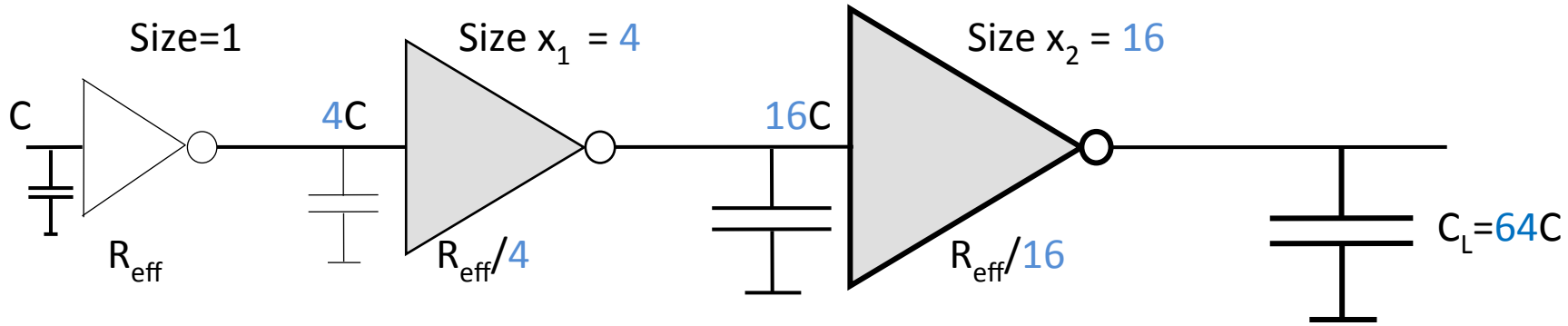
Hint: $\frac{d}{dh_1}D(h_1,h_2)=0$, and $\frac{d}{dh_2}D(h_1,h_2)=0$

yields $h_1 = h_2$ and $h_3 = H/h_1h_2 \rightarrow h = h_1 = h_2 = h_3 = \sqrt[3]{H}$

The tapered buffer with 3 stages

Reference inverter . . . and two inserted buffer inverters

Example with path electrical effort H=64



Normalized path delay: $D = 3\,p_{inv} + h_1 + h_2 + h_3$ where $h_1=x_1$, $h_2=x_2/x_1$ and $h_3=H/x_2$.
But only h_1 and h_2 are independent variables, h_3 becomes $h_3=H/h_1h_2$:

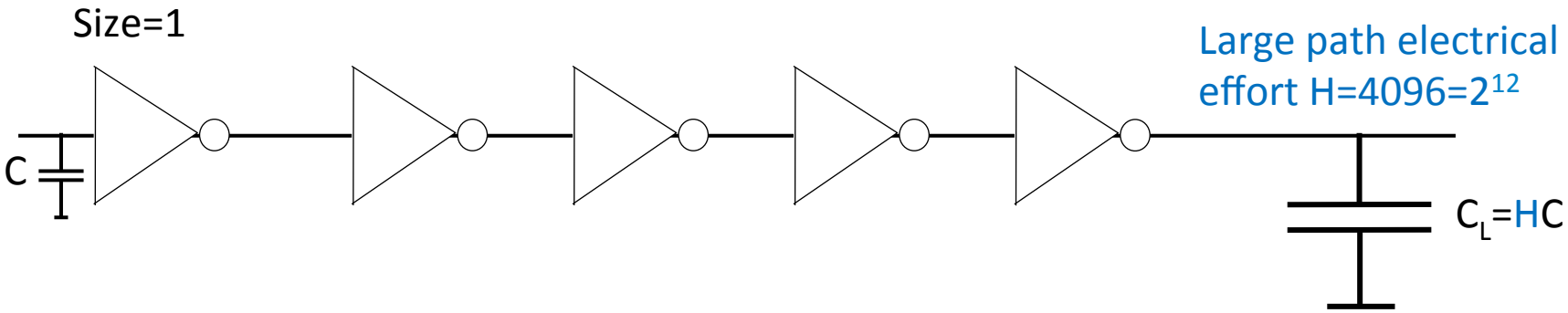
$$D = 3\,p_{inv} + h_1 + h_2 + H/h_1h_2$$

Example: H = 64

Minimum delay is obtained for $h_1=h_2=\sqrt[3]{64}=4$ gives $D = 3(p_{inv}+4)=15$ with $p_{inv}=1$

The tapered buffer with N stages

- What if the path electrical effort, for some reason, is very large, e.g. H=4096.
- How many inverters, *N*, are needed to minimize the delay?



Minimum path delay $D=N(p_{inv}+h)$? Determine best *N* and *h*!

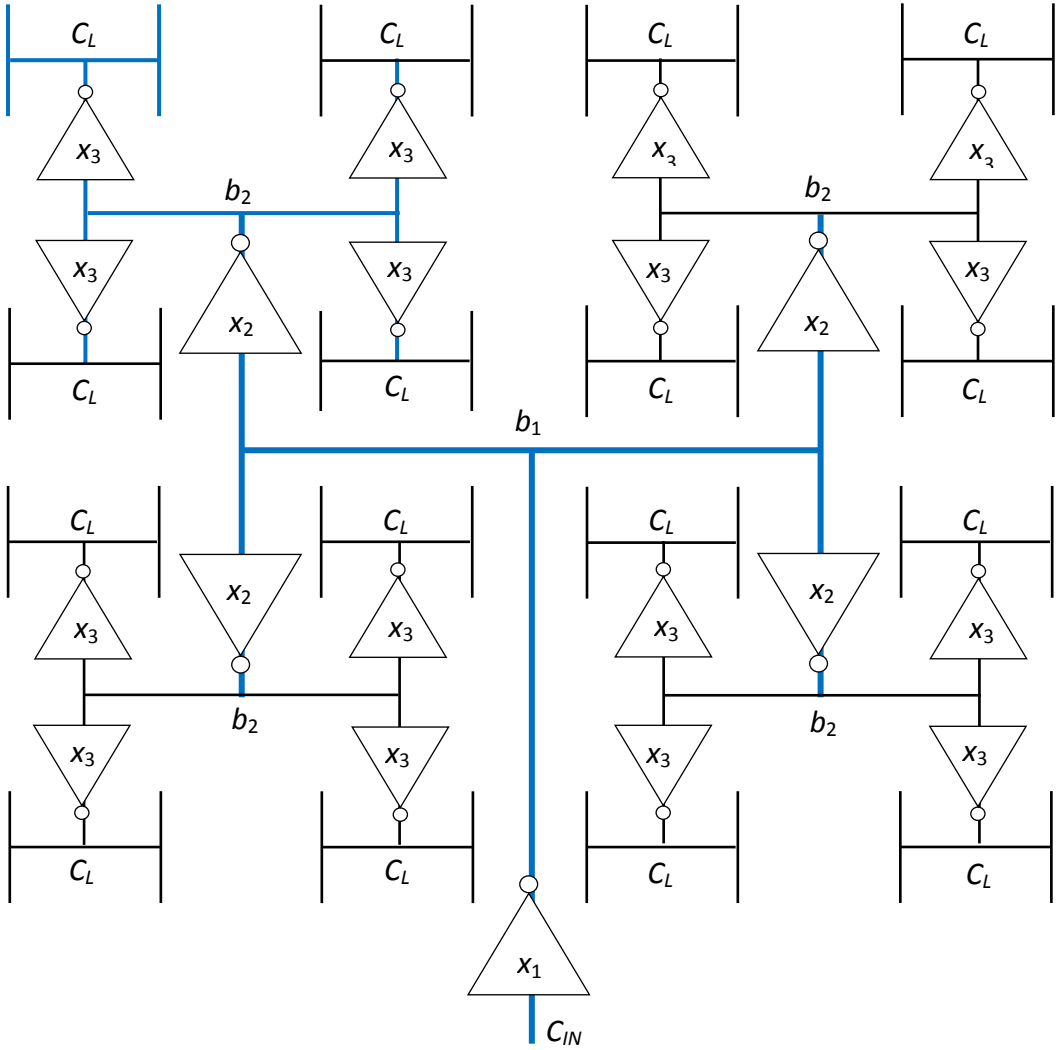
Best h & N?

- Optimal h with p_{inv}=0 can be shown analytically to be h=e= 2.72 for N_{opt} = ln(H)
- With larger p_{inv} numerical solution gives h≈4
  - That is why the FO4 delay is so important!
- The minimum is, so a bit larger h does only increase delay marginally and substancially smaller area

Quick question

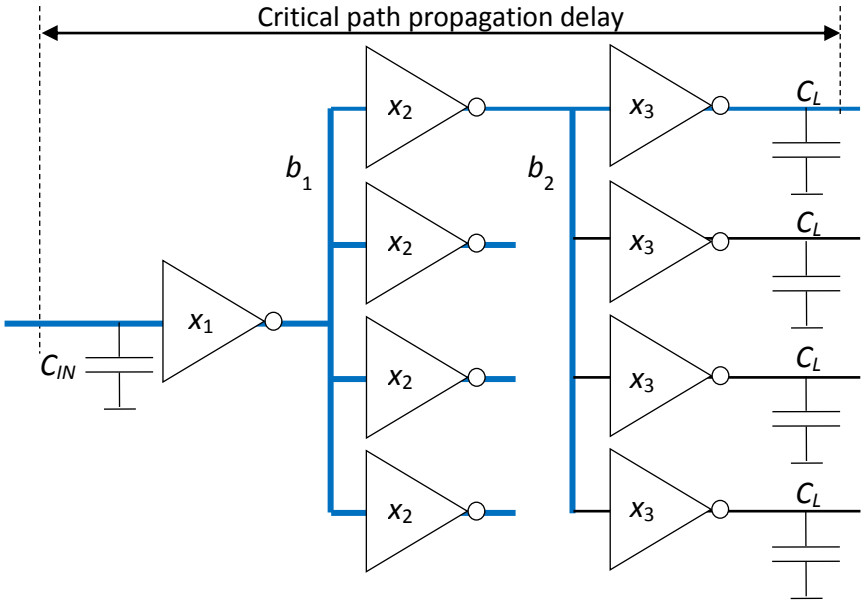
- If you were to design a tapered buffer where the load capacitance is 1500 times larger than the input capacitance, and you are not allowed to invert the signal, how many inverters would you choose?
- What would be the resulting normalized delay for the buffer?

H-tree clock distribution



H-tree clock distribution

- What is the timing path electrical effort?
- What sizes to choose for inverters in the H-tree?



Short summary

The inverter propagation delay, with effective resistance

$$t_{pd} = 0.7 C_L \frac{V_{DD}}{I_{DSAT}} \qquad t_{pd} = 0.7 C_L R_{eff}$$

The technology time constant, tau

$$\tau = 0.7 R_{\text{eff}} C_G$$

The normalized delay

$$d \equiv \frac{t_{pd}}{\tau}$$

Normalized inverter pair delay

$$d = p_{\text{inv}} + 1$$

Normalized fanout-of-four (FO4) delay

$$d = p_{\text{inv}} + 4$$

Tapered buffer with path fanout H and N stages (D is normalized path delay)

$$h_{\text{opt}} = \sqrt[N]{H} \qquad D_{\min} = N(p_{\text{inv}} + h_{\text{opt}})$$

Optimal h is around 4

Long summary

- Defined rise and fall delays at the 50% level ($V_{DD}/2$) andrise and fall times between 20% and 80% levels
- Calculated propagation delay in response to a square-wave input signal assuming MOSFETs being saturated during delay
- Improved the delay model by adding 40%
  - assuming a ramp input signal and
  - assuming equal input and output edge rates
- Replaced saturation current sources by effective resistances
- Made the pMOSFET twice as wide to compensate for lower hole mobility
  - Both MOSFETs now have the same effective resistance of 2 kΩ·μm
  - However, p-channel device now has twice the input capacitance of the n-channel MOSFET
- Obtained an electrical two-port model of the inverter for delay calculations
  - we know what this model looks like seen from the input port, and seen from the output port
- Calculated the FO4 delay, and we found the $R_{eff}C_G$ product being independent of the inverter size (as long as we keep same ratio between W_P and W_N)
- Introduced technology time constant, tau, and normalized delay, d
- Found best inverter sizes of a tapered inverter buffer for minimum delay. Fanout h = 4 is the optimum.