Delay Definitions

- t_{pdr}: rising propagation delay

  From input to rising output crossing $V_{DD}/2$
- t_{pdf}: falling propagation delay

  From input to falling output crossing $V_{DD}/2$
- t_{pd}: average propagation delay

  $t_{pd} = (t_{pdr} + t_{pdf})/2$
- t_r: rise time

  From output crossing 20% to 80% V_{DD}
- t_f: fall time

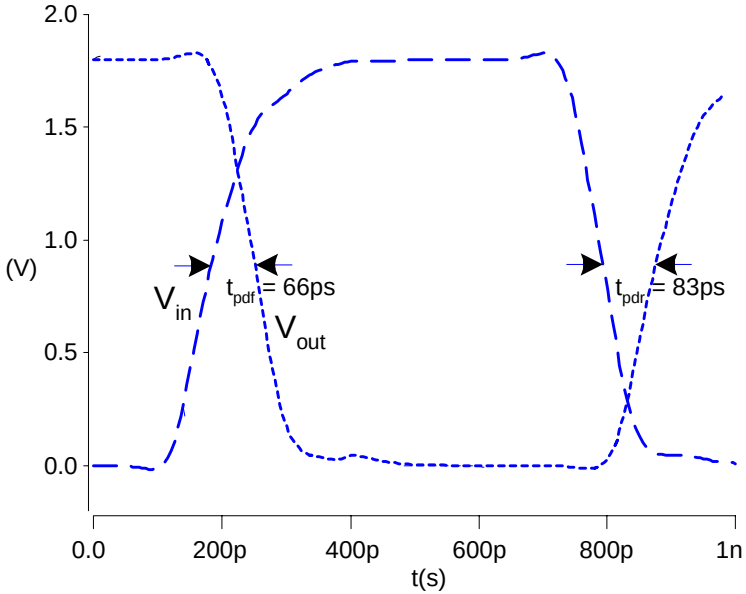
  From output crossing 80% to 20% V_{DD}
- t_{cd}: average contamination delay

  $t_{cd} = (t_{cdr} + t_{cdf})/2$
  - t_{cdr}: rising contamination delay: Min from input to rising output crossing $V_{DD}/2$
  - t_{cdf}: falling contamination delay: Min from input to falling output crossing $V_{DD}/2$

Delay Estimation

Solving differential equations by hand is hard.

SPICE like simulators used for accurate analysis. But simulations are expensive.



We need to be able to estimate delay although not as accurately as simulator.

Use RC delay models to estimate delay

- C = total capacitance on the output node
- Use Effective resistance R
- Therefore t_{pd} = RC

Transistors are characterized by finding their effective R.

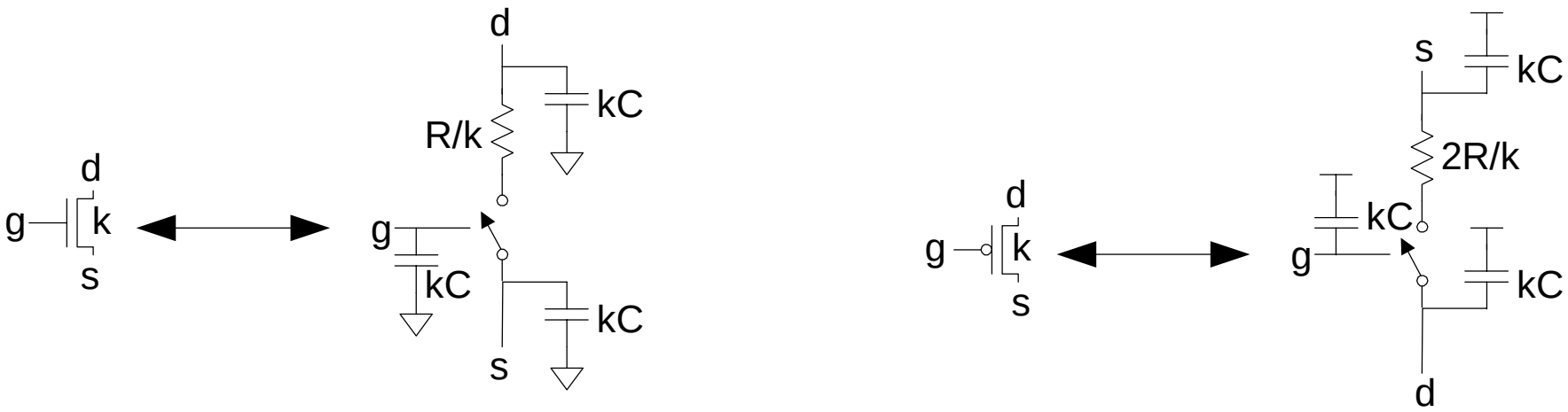
RC Delay Models

Equivalent circuits used for MOS transistors

- Ideal switch + capacitance and ON resistance
- Unit NMOS has resistance R, capacitance C
- Unit PMOS has resistance 2R, capacitance C

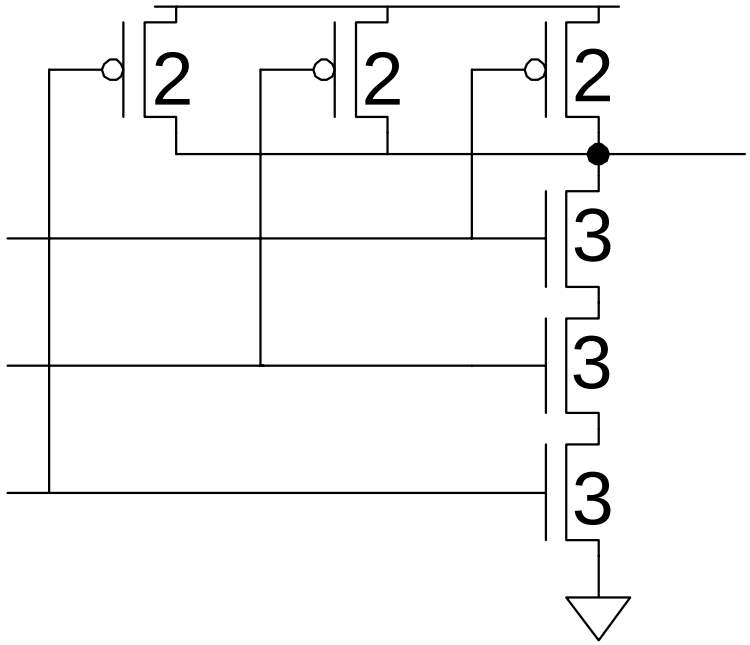
Capacitance proportional to width

Resistance is inversely proportional to width



RC Delay Models

A 3-input NAND gate with transistor widths chosen to achieve effective rise and fall resistance equal to that of a unit inverter (R)



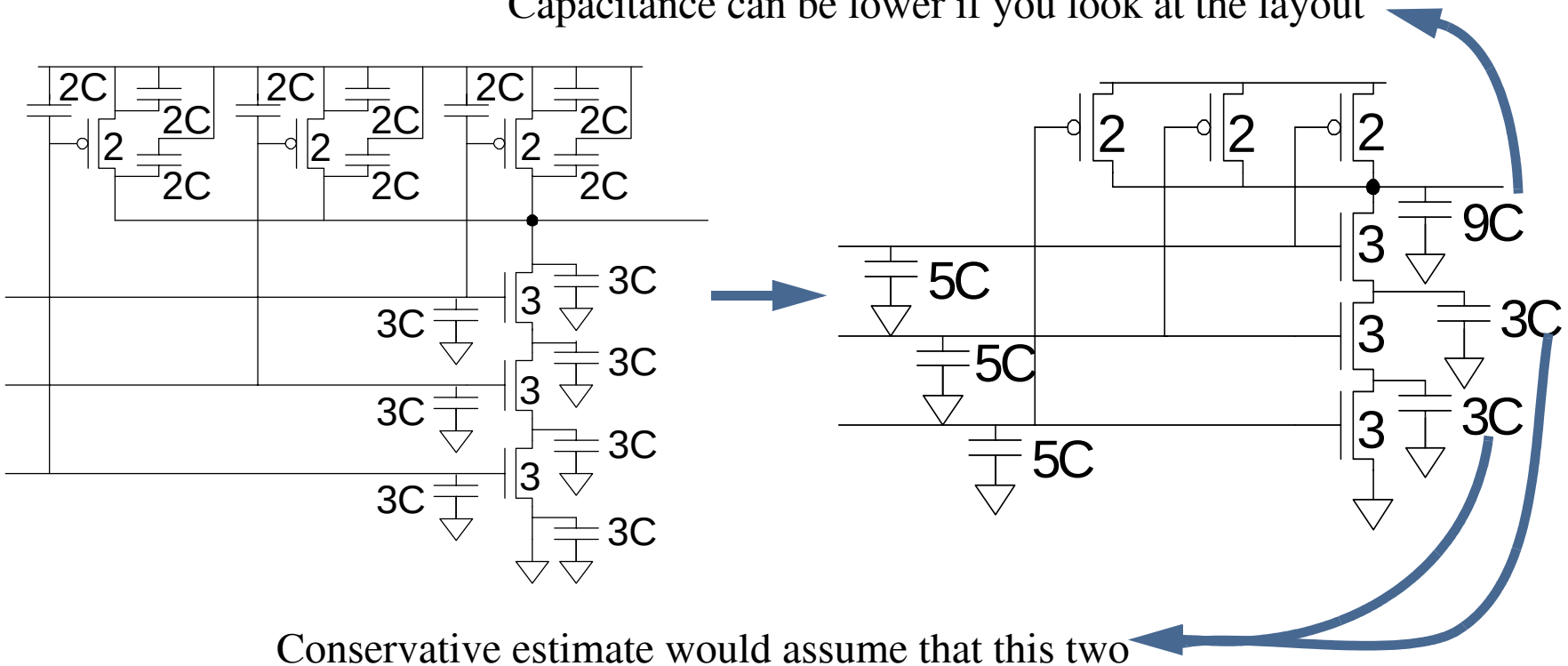
3 NMOS in series = 3R, therefore each has to be three times the minimum width.

3 PMOS(2R) in parallel, worst case one ON, therefore each has to be twice minimum width

RC Delay Models

3-input NAND gate with it's gate and diffusion capacitances (assuming all nodes are contacted). Estimated at the schematic level, values will be different if you look at layouts.

Capacitance can be lower if you look at the layout

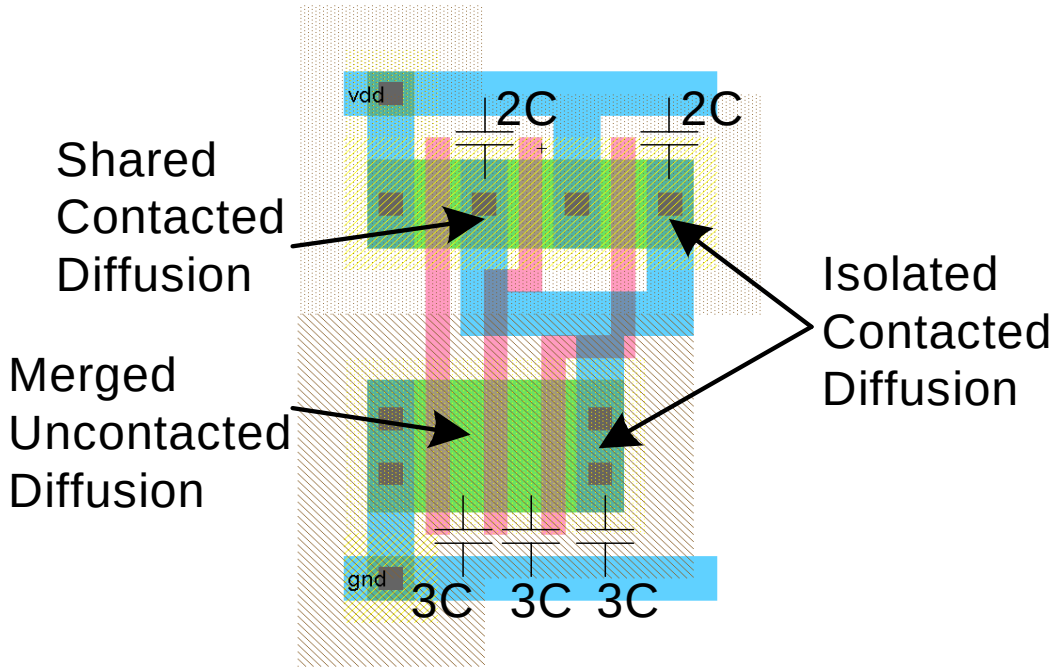


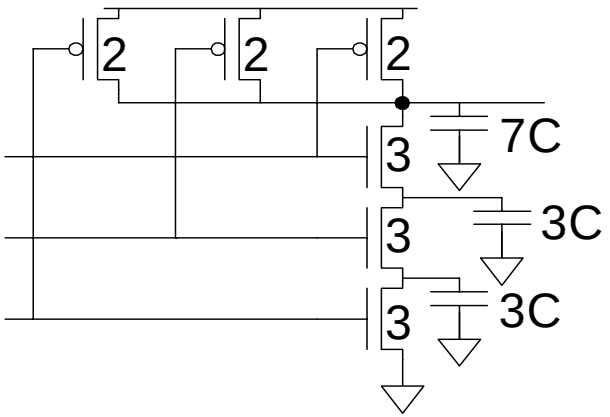
Conservative estimate would assume that this two diffusions are uncontacted and therefore have lower capacitance (the difference is usually ignored for hand calculations)

RC Delay Models: Layouts

Good layout minimizes diffusion area.

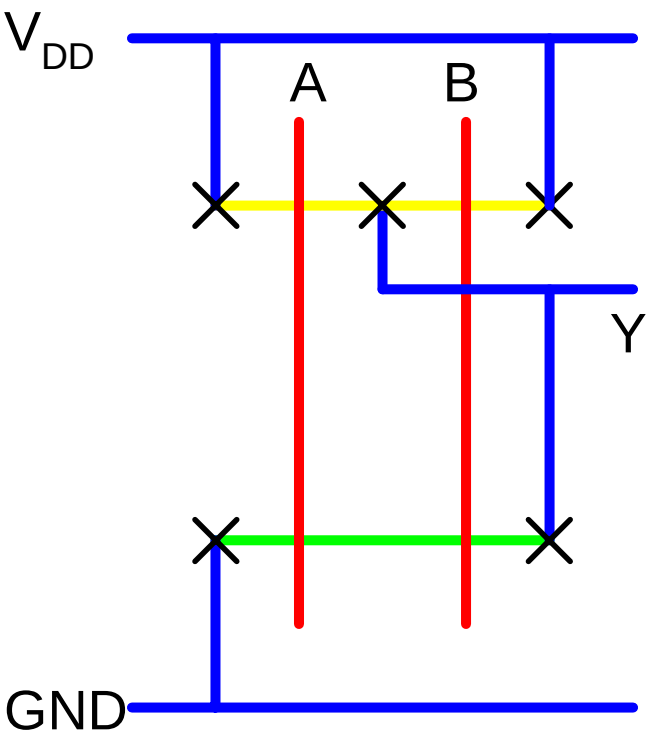
NAND3 gate shown below, shares one diffusion contact, thus lowering the output capacitance by 2C. Contacted diffusions are assumed.

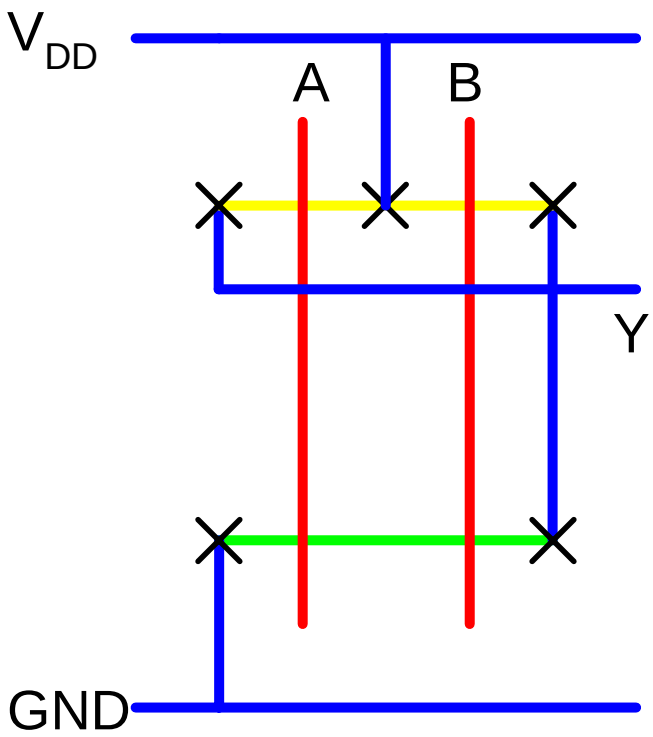




RC Delay Models: Layout Comparison

Which layout is better?



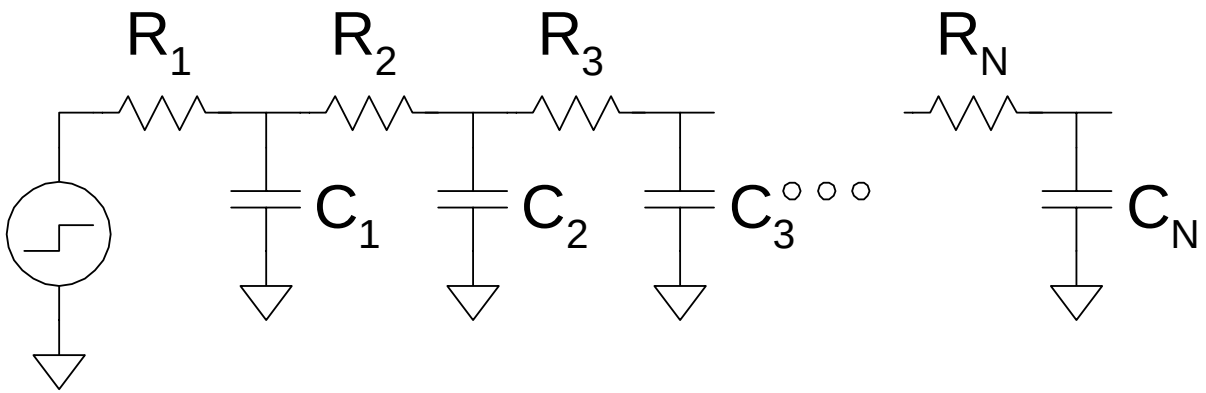


Elmore Delay

ON transistors are considered as resistors.
Pull-up or pull-down networks are considered as RC ladders.
Elmore Delay of a RC ladder:

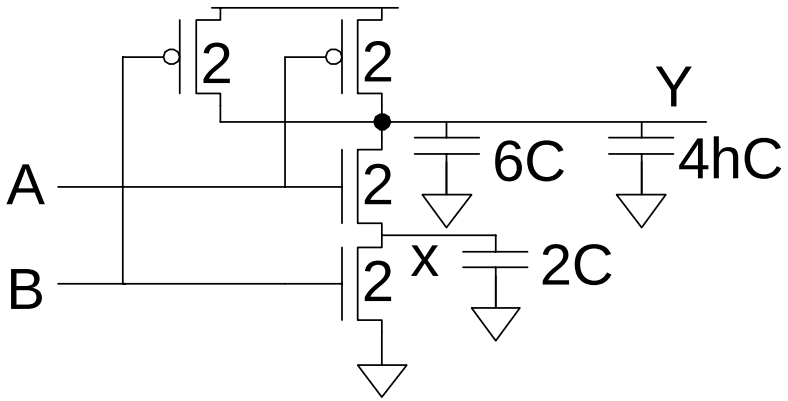
$$t_{pd} = \sum_{nodes\ i} R_{n-i} C_i$$

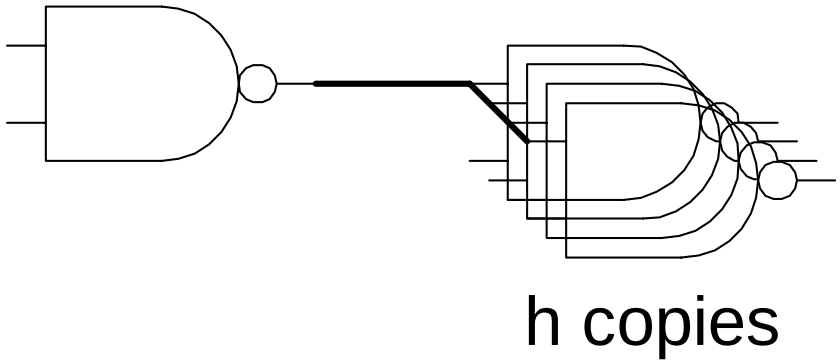
$$= R_1 C_1 + (R_1 + R_2) C_2 + \ldots + (R_1 + R_2 + \ldots + R_N) C_N$$

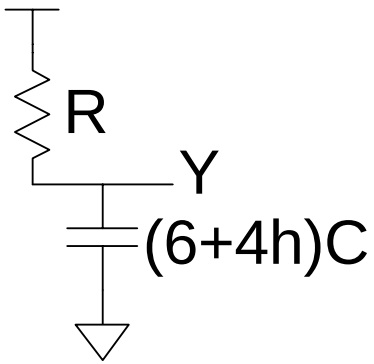


Elmore Delay

Rising propagation delay of a 2-input NAND gate driving *h* identical NAND gates



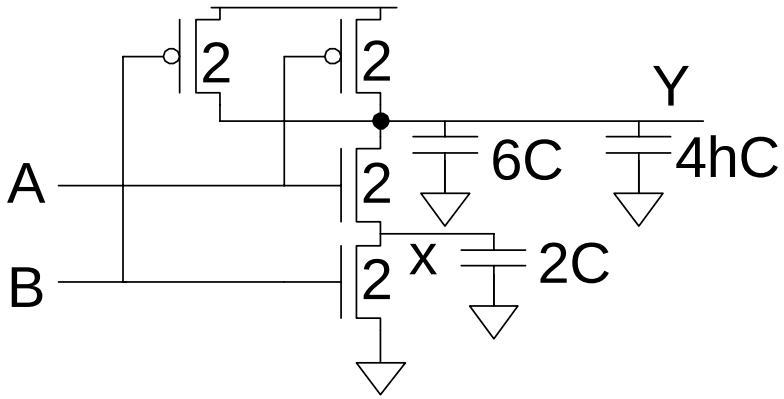


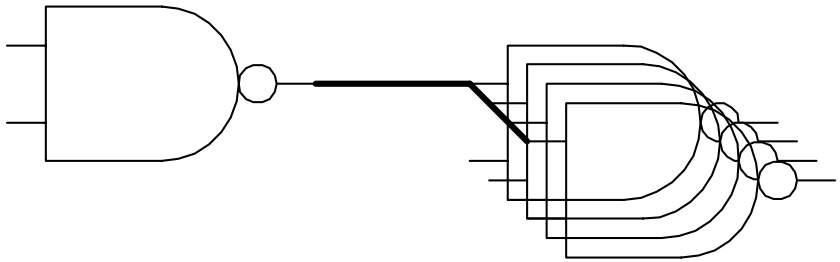


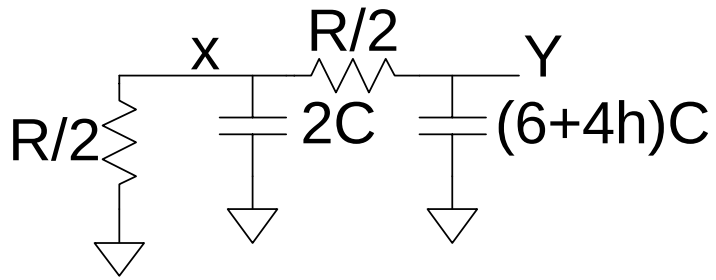
$$t_{pdr} = (6 + 4h)RC$$

Elmore Delay

Falling propagation delay of a 2-input NAND gate driving *h* identical NAND gates







$$t_{pdf} = (2C)\left(\frac{R}{2}\right) + [(6+4h)C]\left(\frac{R}{2}+\frac{R}{2}\right)$$

$$= (7+4h)RC$$

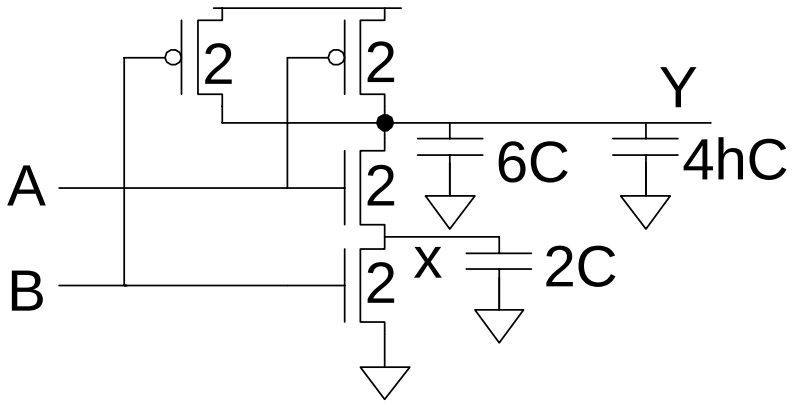
Delay Components and Contamination Delay

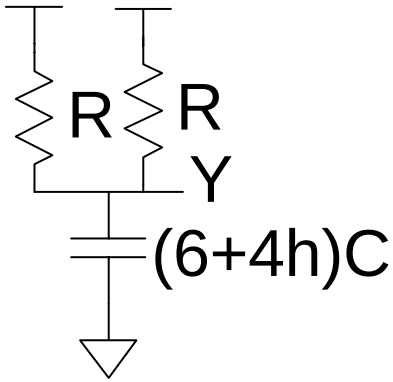
Total delay is composed of two parts:

- *Parasitic delay*: 6 or 7 RC in previous example, independent of load
- *Effort delay*: 4*h* RC in previous example proportional to load capacitance.

Contamination Delay (best case delay): can be substantially less than propagation delay.

Example: Both inputs fall simultaneously in 2-input NAND gate.





$$t_{cdr} = (3 + 2h)RC$$