Chapter 3

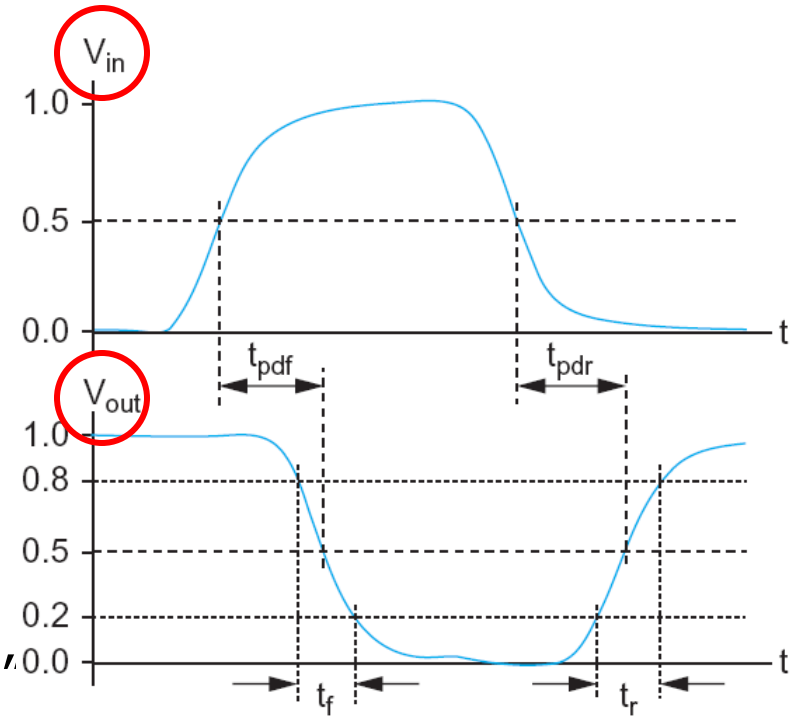
Speed (Delay)

闕志達

台灣大學電機系

Delay in Circuits

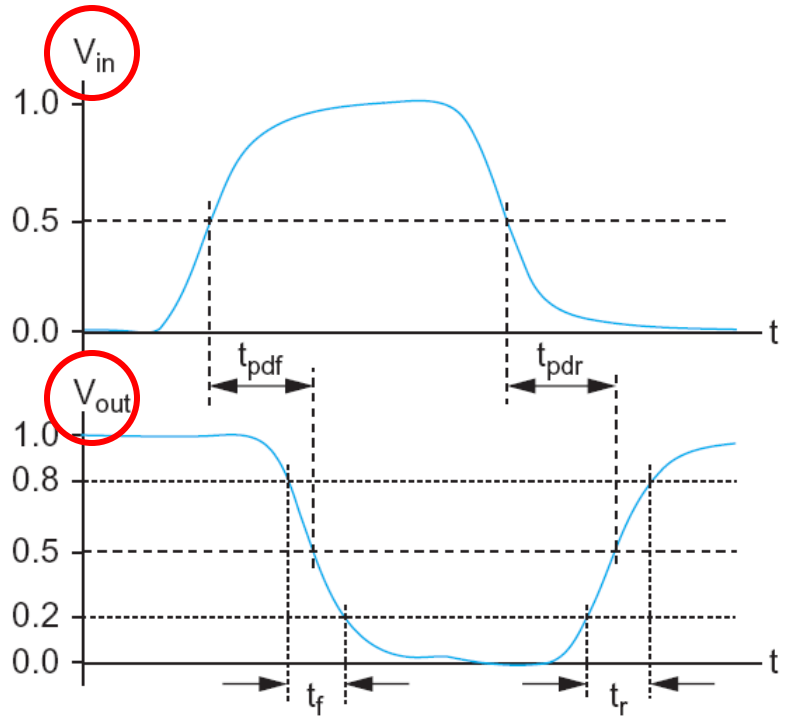
- Propagation delay time, t_{pd} = maximum time between input transition (50%) to output crossing 50% output level.
  - t_{pdf} , t_{pdr}
- Contamination delay time, t_{cd} = minimum time between input transition (50%) to output crossing 50% output level.
- When a signal travels with contamination delay and violate hold time of the following FF, the output value may be "contaminated" by the faster-than-normal signal. This condition is called "race."



Transition Times

- Rise time, t_r = time for a waveform to rise from 20% to 80% of its steady-state value.
- Fall time, t_f = time for a waveform to fall from 80% to 20% of its steady-state value.
- Average transition time,

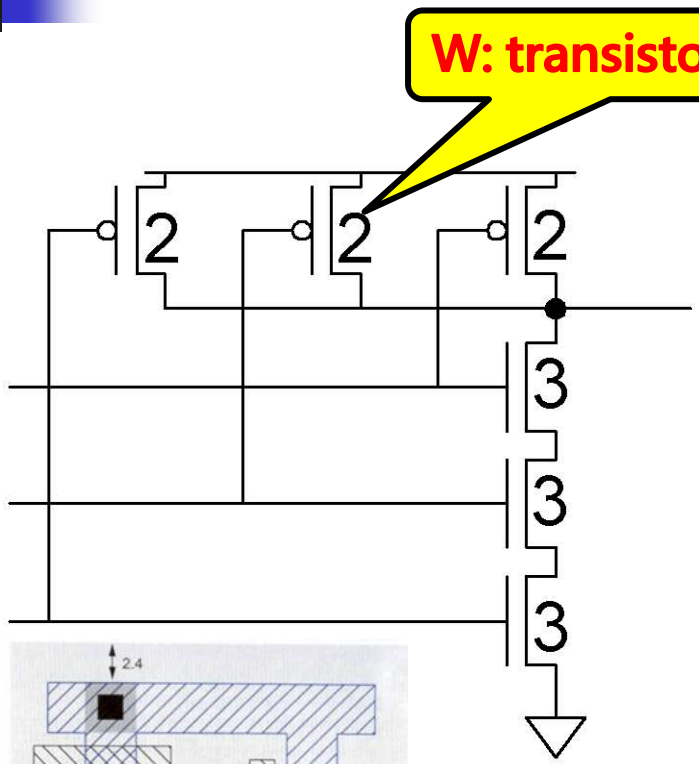
$$t_{rf} = (t_r + t_f)/2$$



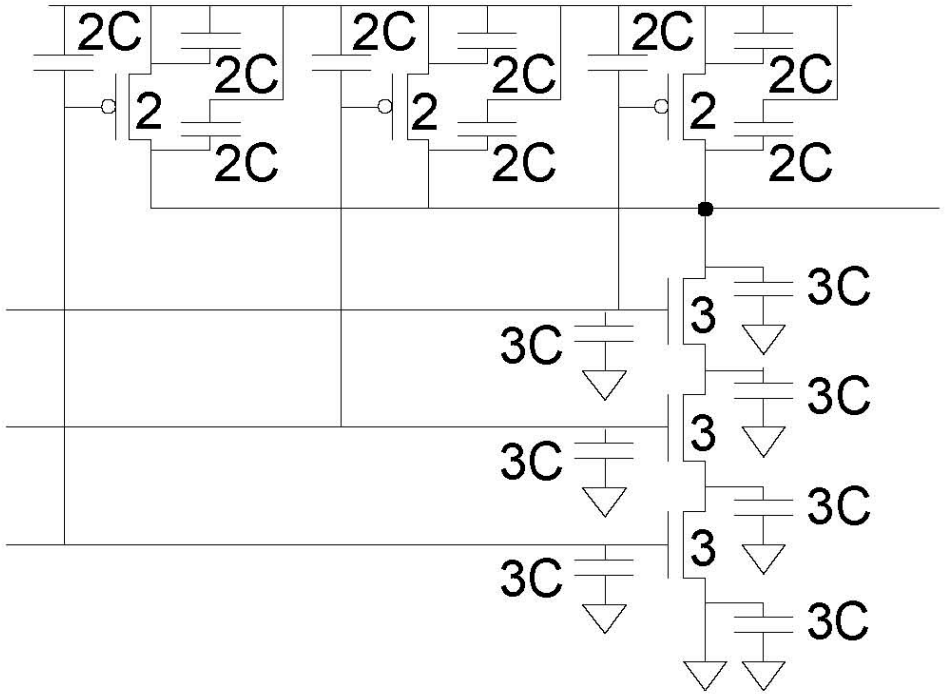
Gate RC Delay Estimation

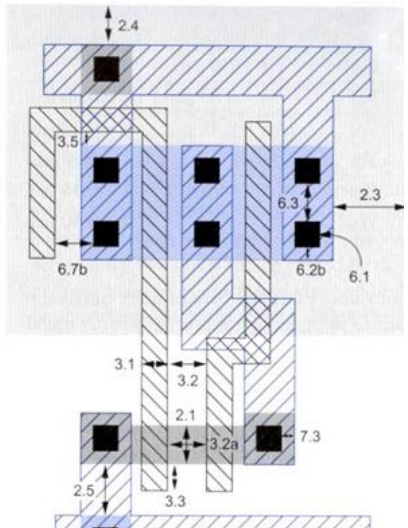
- Delay of a gate is inversely proportional to its driving strength (1/R) and proportional to its load capacitance (C).
- Note that we can assume PMOS needs be twice as wide as NMOS to get same driving strength. Assume $u_n = 2u_p$
- Resistor
  - ON transistor can be modeled as a resistor.
- Capacitance
  - Gate capacitance
  - Diffusion capacitance
  - Wire capacitance

RC Analysis Example

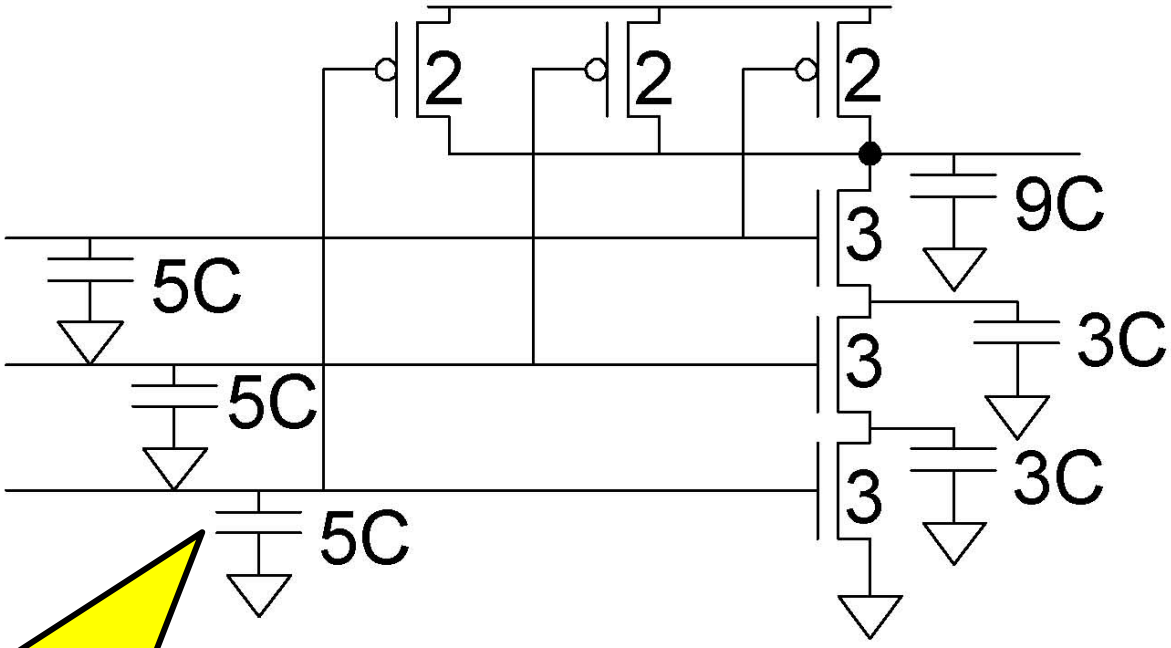


All transistors have same L





Equivalent Circuit with RC



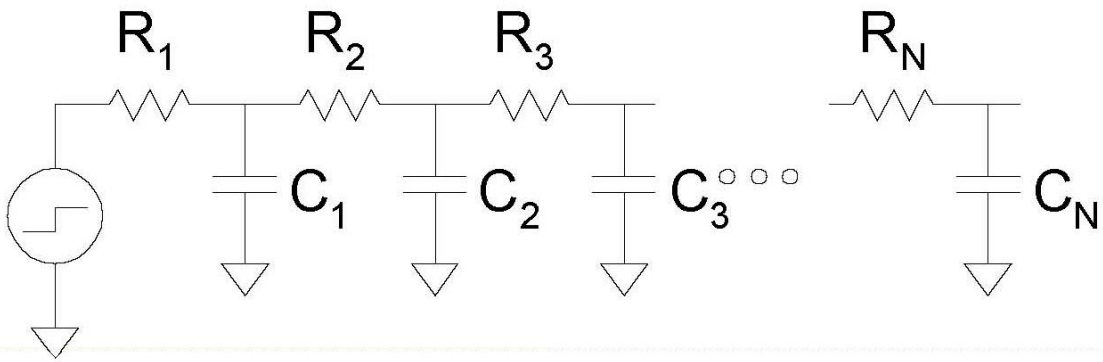
Treat Vdd as AC GND

Elmore Delay Model

- Pull-up and pull-down network are like RC ladder
- R is on the shared path from the source to the node and the leaf (the node to compute delay).

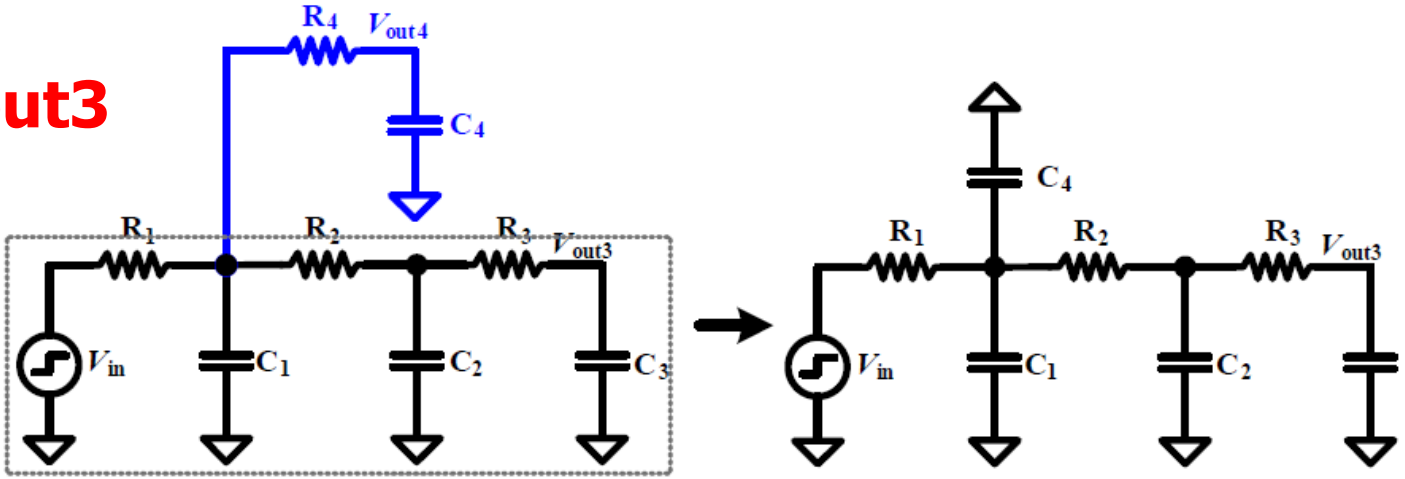
$$t_{pd} \approx \sum_{\text{nodes } i} R_{i-to-source} C_i$$

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

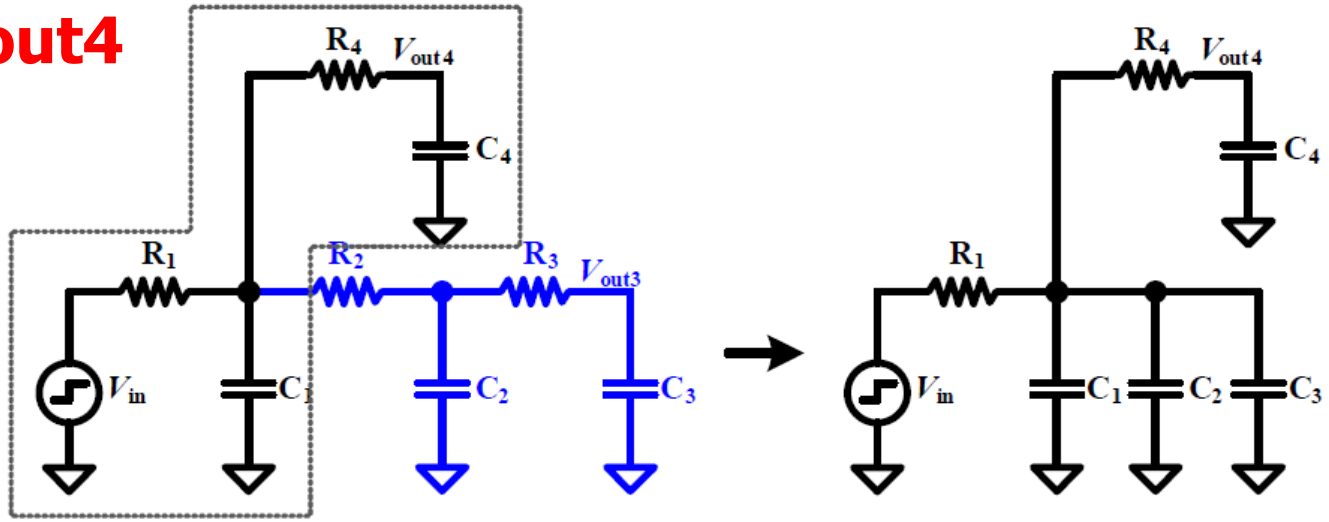


Elmore Delay Model Examples

- **Compute Vout3**



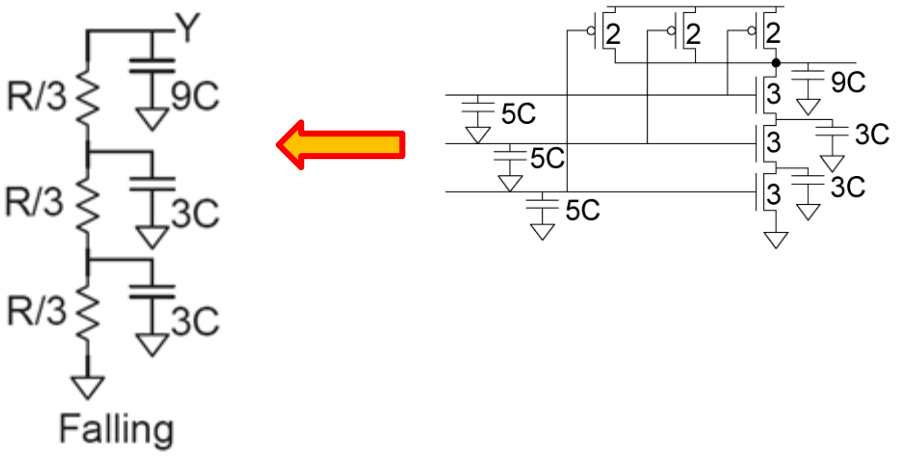
- **Compute Vout4**



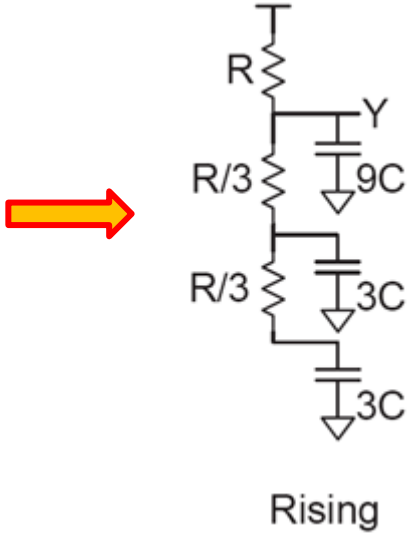
Delay Analysis Example

- 3-input NAND, no load

When Y goes LOW



When Y goes HIGH

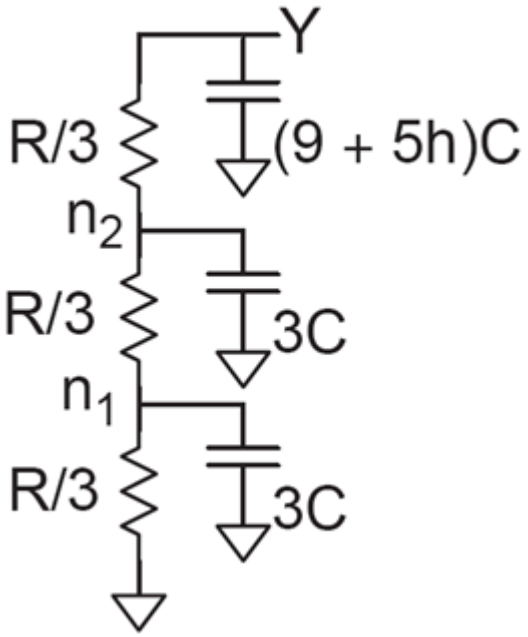


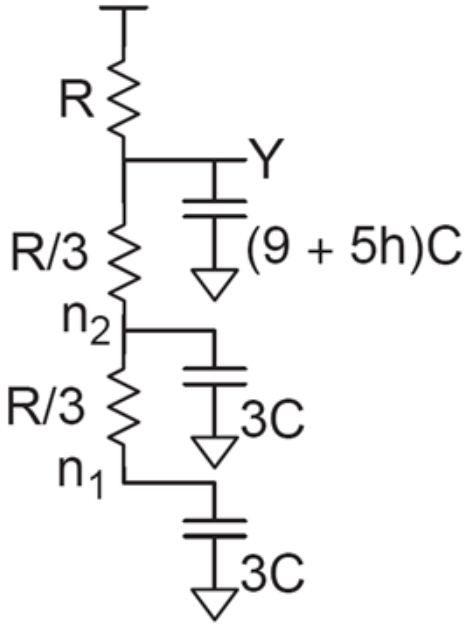
$$t_{pdf} = R(9C) + (2R/3)(3C) + (R/3)(3C)$$
$$= 12RC$$

$$t_{pdr} = R(9C) + R\,(3C) + R(3C)$$
$$= 15RC$$

Loaded Analysis

- *h* is the number of fanout 3-input NAND gates





$$t_{pdr} = R(9C + 5hC) + R\,(3C) + R\,(3C)$$
$$= \underline{(15 + 5h)RC}$$

$$t_{pdf} = R(9C + 5hC) + (2R/3)(3C) + (R/3)(3C)$$
$$= \underline{(12 + 5h)RC}$$

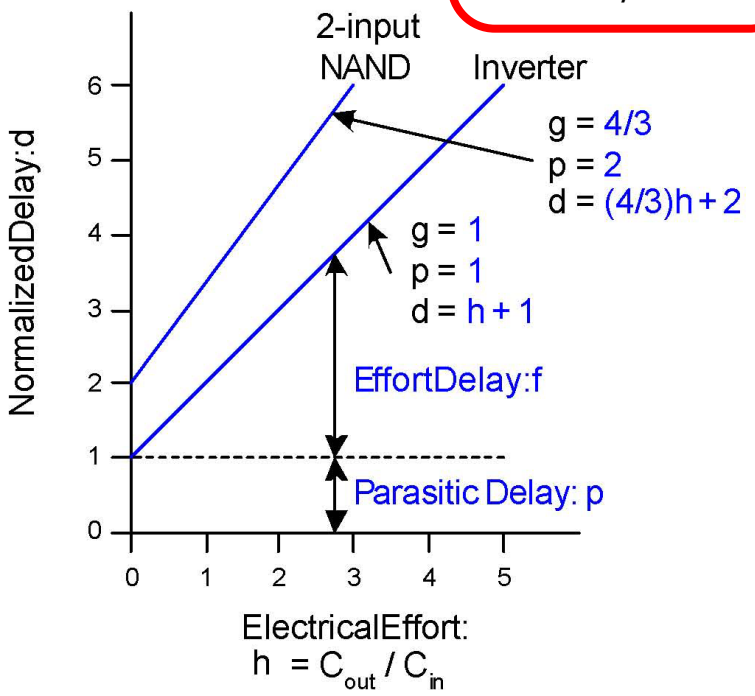
Linear Delay Model

- A common delay estimation approach in IC design is to treat logic gates as simple delay elements with the following delay time

$$t_{pd} = t_{internal} + t_{LE} * h = p + g * h$$

$p = t_{internal}$
$g = t_{LE}$
$d = t_{pd}$

where $t_{internal}$ is the fixed parasitic internal delay and t_{LE} is the unit-output-loading logical effort delay; h is the number of fanout gates of the same type. For different type of gates, $h = C_L/Cin = Cout/Cin$.

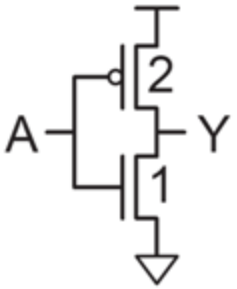


Parasitic Delay ($t_{internal}$)

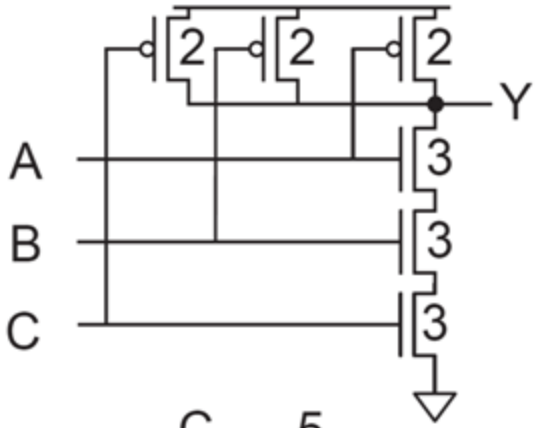
- Delay when the output load is zero. Can be obtained by simulation.
- Approximate formula says that the parasitic delay is proportional to *N*, the number of inputs in NAND/NOR gates.
- In practice, this is not true if we use a more accurate Elmore delay model. In this case, the parasitic delay grows with N^2.
- It is not advisable to use more than 4-5 NMOS/PMOS series transistors.
- NMOS in series is always preferred to PMOS in series.

Logical Effort (g or t_{LE})

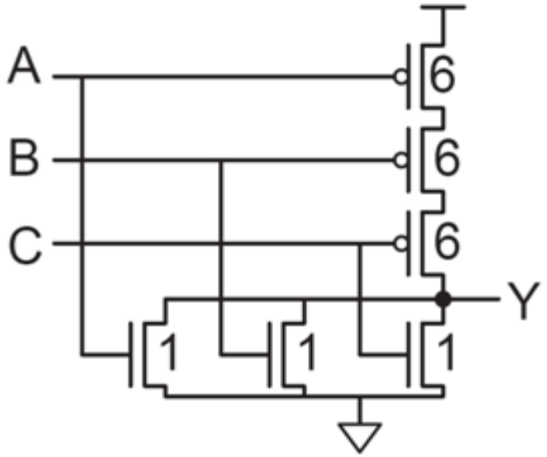
- The ratio of the input capacitance between the current gate and the inverter that can deliver the same output current level.
- Examples (all with unit resistance)



$C_{in} = 3$
$g = 3/3$



$C_{in} = 5$
$g = 5/3$



$C_{in} = 7$
$g = 7/3$

Input Waveform Slope

- The largest source of error in the linear model is the slope of the input waveform.
- The delay time formulas can be modified

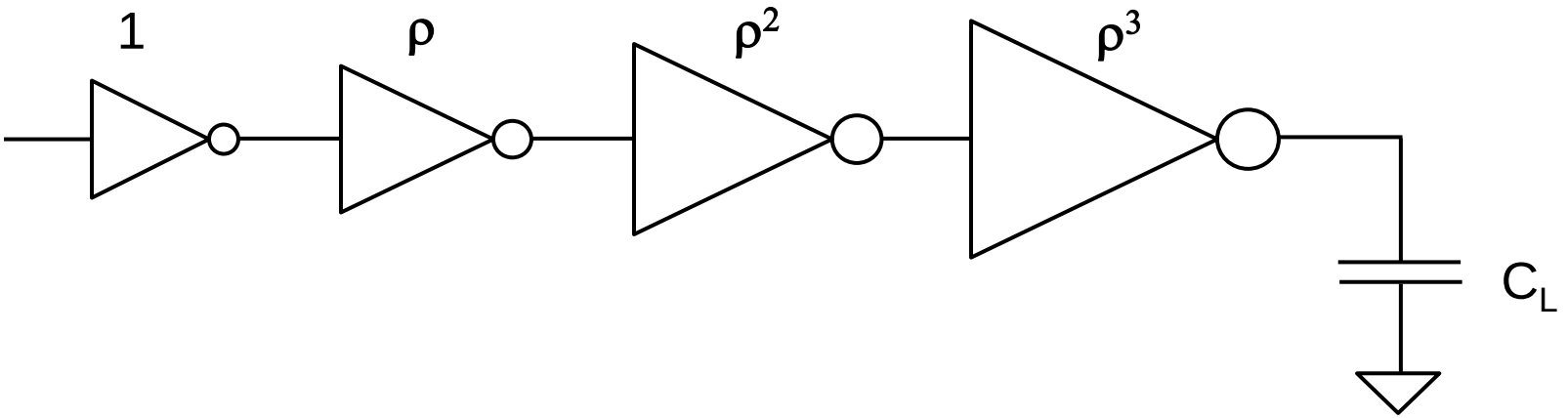
$$t_{pd} = t_{pd\text{-}step} + t_{edge} * k$$

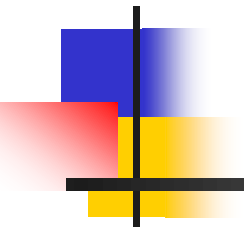
where $t_{pd\text{-}step}$ is the propagation delay assuming step input and t_{edge} is the input rise/fall time.

- Other sources of error in linear delay model include different input arrival times and parasitic capacitance Cgs and Cgd, and wire capacitance.

Stage Ratio

- Very often, large drivers are needed to transmit signals over a long bus or a large load (e.g., output pads and bus driver). This is often achieved by a chain of successively large inverters.
- Usually, we set this ratio to 4.

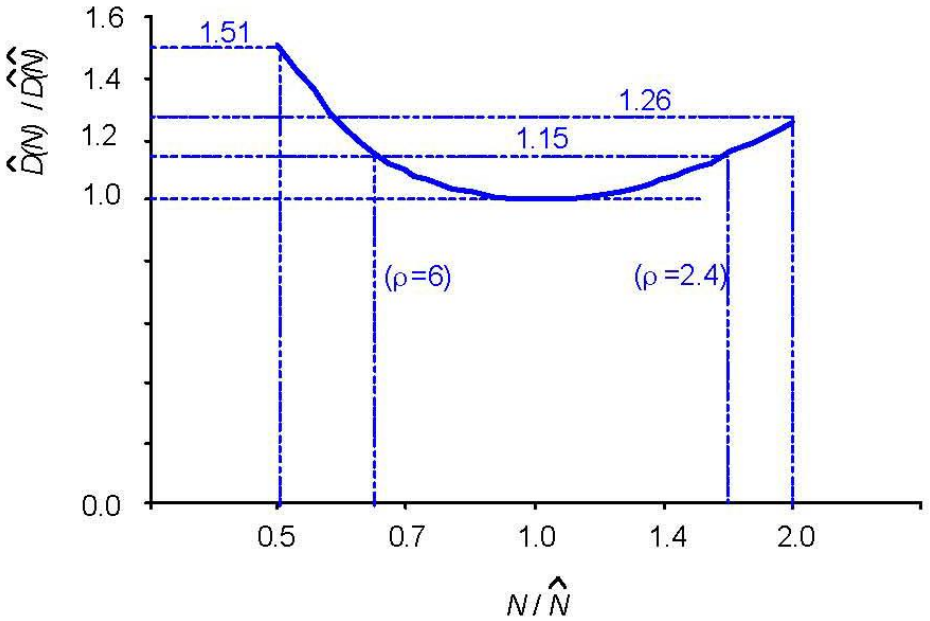




補充教材

Optimal Stage Ratio

- The optimal stage ratio (in the sense of least delay) satisfies the equation: $p_{inv} + \rho(1-\ln\rho) = 0$, where p_{inv} is the parasitic delay and ρ is the stage ratio.
- The optimal is around 3.59 if $p_{inv} = 1$. For convenience 4 is often used. Actually any ρ between 2.4 and 6 gives a delay that is within 15% of the optimal value.



Multistage Path Delay

- Example: Chain of gates. Note *h* = Cout/Cin.

$$t_{pd\text{-}path} = p_i + g_i * h_i$$
$$= [1 + (x/10)] + [2 + (5y/3x)] + [2 + 4z/3y)]$$
$$+ [1 + 20/z]$$

- p ($t_{internal}$) of inverter is 1, and p of 2-input NAND/NOR is 2. See Table 3.3 of textbook.