EECS 427
Lecture 7: Logical Effort

Reading: handout

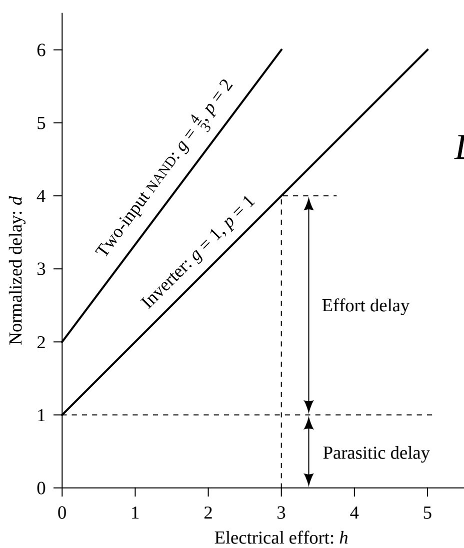
Last Time

- Baseline project architecture
- Intro to logical effort
  - g: Logical effort
  - h: Electrical effort
  - p: Parasitic delay

Outline

- Delay of a chain of gates (multistage)
- Branching
- Minimum delay
- Best number of stages and gate sizing
- Examples
- Limitations

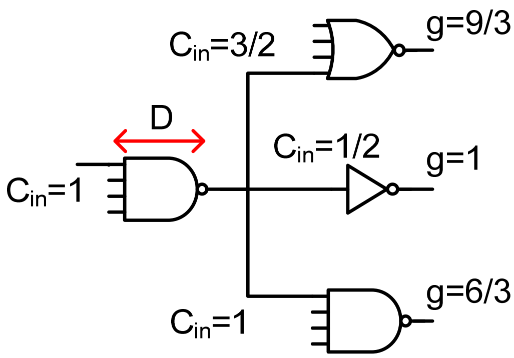
Delay components review



$$D = p + g \cdot h$$

I. Sutherland, *et al*, Logical Effort, Academic Press, 1999

More complex circuit



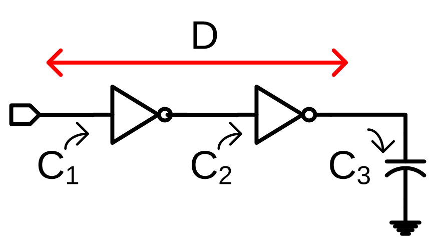
$$h = 3$$

$$g = \frac{6}{3}$$

$$p = 4 \times 0.6$$

$$D = 8.4\tau$$

Multistage effort



Define path effort H as:

$$H = \frac{C_3}{C_1}$$

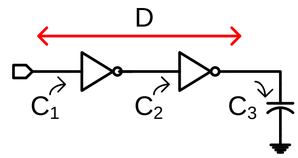
$$h_1 = \frac{C_2}{C_1}$$

$$h_2 = \frac{C_3}{C_2}$$

$$\Rightarrow \quad H = h_1 h_2$$

Minimum delay

Total delay $D = g_1 h_1 + p_1 + g_2 h_2 + p_2$



Substitute H $D = g_1 h_1 + p_1 + g_2 \frac{H}{h_1} + p_2$

$$\min\{D\} \Rightarrow \frac{\partial D}{\partial h_1} = g_1 - g_2 \frac{H}{h_1^{\,2}} = 0$$

Minimize the delay

$$g_1 - g_2 \frac{h_2}{h_1} = 0$$

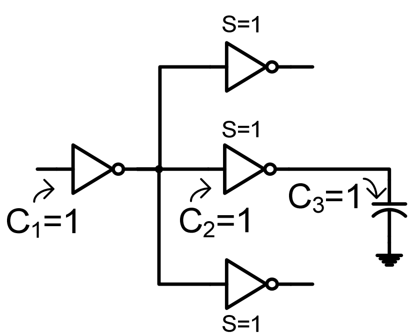
Solve for the minimum delay

$$g_1 h_1 = g_2 h_2$$

$$f_1 = f_2$$

Delay is optimal when stage efforts are equal

Branching



Without branching: $H = h_1 h_2$

With branching: $h_2 = \frac{C_3}{C_2}$

$$h_1 = \frac{3C_2}{C_1}$$

$$h_1 h_2 = 3H$$

$$b_i = \frac{C_{on-path} + C_{off-path}}{C_{on-path}}$$

$$C_{on-path,2} = C_2$$

$$C_{off-path,2} = 2C_2$$

$$b_2 = \frac{3C_2}{C_2} = 3$$

Equivalent Path Efforts

$$H = \frac{C_{out}}{C_{in}}$$

$$B = \prod b_i$$

$$\prod h_i = BH$$

$$G = \prod g_i$$

Path Effort

$$F = GBH = \prod g_i h_i$$

Path Effort:

- Does not change with added inverters
- Does not depend on sizes, but on topology

Minimum Delay

$$F = GBH = \prod g_i h_i$$

Stage effort of stage i: f_i

$$g_i h_i = f_i$$

Optimal stage effort is:

$$\hat{f} = f_i$$

For N stages:

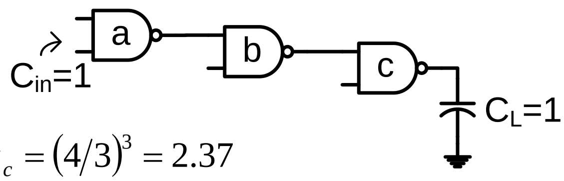
$$F = \hat{f}^{\,N} \Rightarrow \boxed{\hat{f} = F^{1/N}}$$

Minimum Delay:

$$\hat{D} = \sum_i g_i h_i + p_i$$

$$\hat{D} = N\hat{f} + \sum_i p_i = \boxed{NF^{1/N} + P}$$

Example to compute min. delay



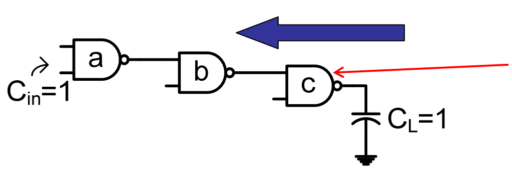
$$G = g_a g_b g_c = (4/3)^3 = 2.37$$
$$B = 1$$
$$H = C_L / C_{in} = 1$$
$$F = GBH = 2.37$$
$$\hat{f} = F^{1/3} = 4/3$$
$$\hat{D} = 3(2.37)^{1/3} + 3(2p_{inv})$$
$$\hat{D} = 4 + 6 \times 0.6 = 7.6\tau$$

Stage sizing

After computing $\hat{f}$

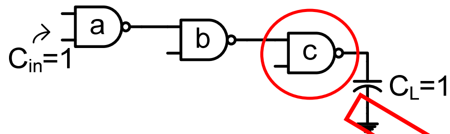
$$h_i = \hat{f} / g_i = C_{out} / C_{in}$$
$$C_{in} = \frac{g_i C_{out}}{\hat{f}}$$

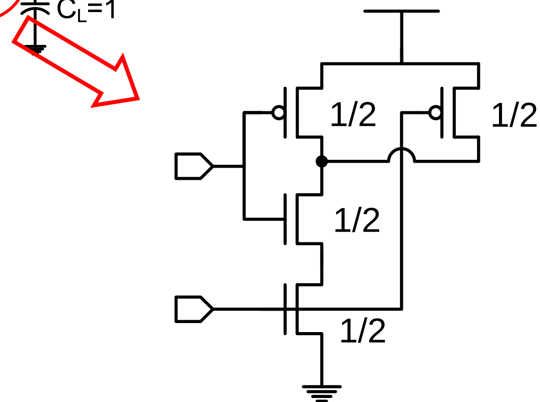
Work backwards to size each gate



$$C_{in,c} = \frac{g_{nand} C_L}{\hat{f}}$$

Stage sizing example





$$C_{in,c} = \frac{g_{nand} C_L}{\hat{f}}$$

$$C_{in,c} = \frac{4/3C}{4/3} = C$$

Calculating optimal # of stages

- Path effort F can be used to determine the optimal number of stages
  - Assuming we add n_2 inverters
    - New number of stages $N=n_1+n_2$
    - G, B, H don't change – F is fixed
    - But P increases

$$\hat{D} = NF^{1/N} + \sum_{i}^{n_1} p_i + n_2 p_{inv}$$

Optimum is technology dependent

Best number of stages for $p_{inv} = 0.6\tau$

Path effort F	Best number of stages $\hat{f}$	Min. delay $\hat{D}$	Stage effort f
0		1.0	
	1		0-5.1
5.13		5.7	
	2		2.3-4.2
17.7		9.6	
	3		2.6-3.9
59.4		13.5	
	4		2.8-3.7
196		17.4	
	5		2.9-3.6
647		21.2	
	6		2.9-3.6
2130		25.1	
	7		3.0-3.5
6980		29.0	

I. Sutherland, *et al*, Logical Effort, Academic Press, 1999

Summary of the terminology

Gate level

Parasitic delay p

Logical effort g

Electrical effort $h = \frac{C_{in}}{C_{out}}$

Stage effort $f = gh$

Stage delay $d = f + p$

Path

Path electrical effort $H = \frac{C_{out-path}}{C_{in-path}}$

Path logical effort $G = \prod g_i$

Branch effort $B = \prod b_i$

$\prod h_i = BH$

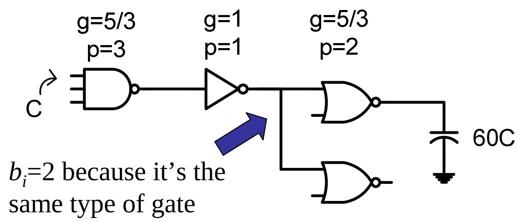
Branching factor $b_i = \frac{C_{on-path} + C_{off-path}}{C_{on-path}}$

Method

Path effort	$F = GBH$
Optimal stage effort	$\hat{f} = F^{1/N}$
Optimal path delay	$\hat{D} = NF^{1/N} + P$
Stage sizing	$C_{in} = \dfrac{g_i C_{out-stage}}{\hat{f}}$

1. Compute path effort
2. Compute optimal stage effort
3. Add buffers (determine optimal number of stages)
4. Size individual gates (working backwards)

Method example 1/3



$$G = \frac{5}{3}\cdot 1\cdot \frac{5}{3} = \frac{25}{9}$$

$$B = 1\cdot 2\cdot 1 = 2$$

$$H = 60$$

$$\sum p_i = 3+1+2 = 6$$

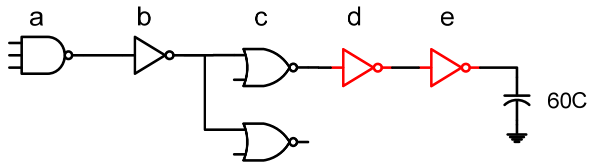
$$F = GBH = 333.33 \rightarrow \hat{N} = 5$$

$$\hat{D} = 5(333.33)^{1/5} + 6 + 2 = 24\tau$$

$$\hat{f} = (333.33)^{1/5} = 3.2$$

- From the table, the optimal # of stages for F=333.33 is 5
- Add 2 inverters to the existing 3 stages

Method example: Sizing 2/3



$$C_{in,e} = \frac{g_i C_{out,e}}{\hat{f}} = \frac{1 \times 60C}{3.2} = 18.75C$$

$$C_{in,d} = \frac{1 \times 18.75C}{3.2} = 5.86C$$

$$C_{in,c} = \frac{5/3 \times 5.86C}{3.2} = 3.05C$$

$$C_{in,b} = \frac{1 \times 3.05C \times 2}{3.2} = 1.96C$$

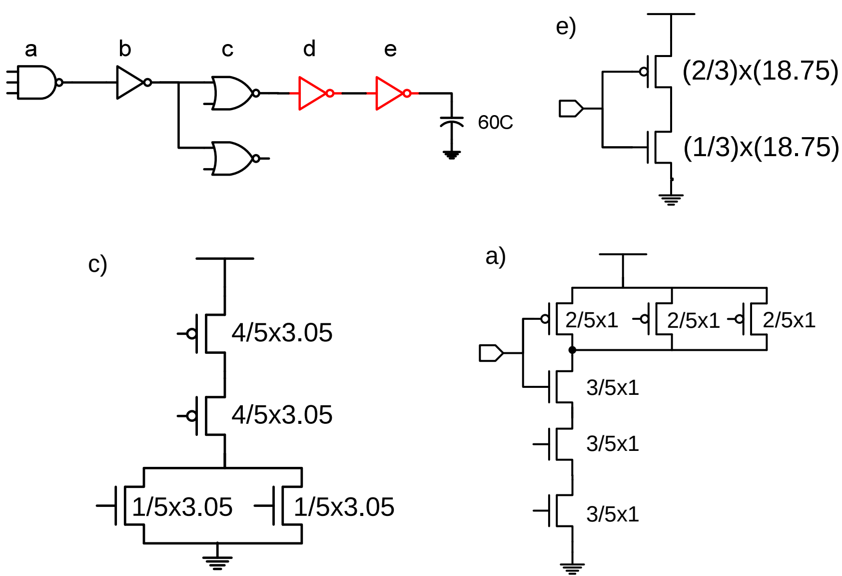
$$C_{in,a} = \frac{5/3 \times 1.9C}{3.2} = 0.989C \approx C$$

← Always check this

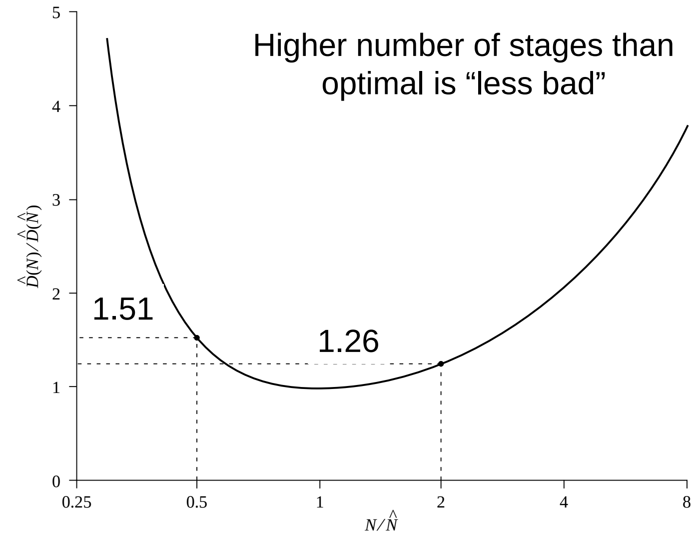
When there is a branch:

$$C_{in} = \frac{g_i C_{out} b_i}{\hat{f}}$$

Method example: Gate sizing 3/3

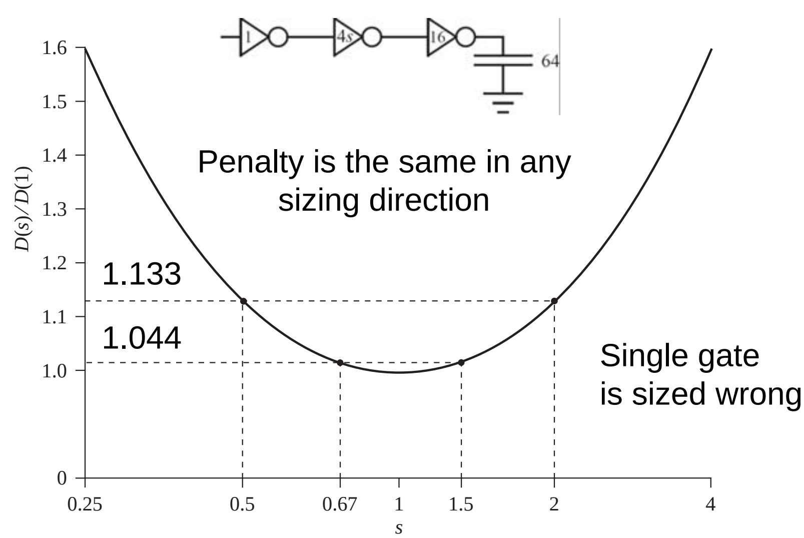


Wrong number of stages



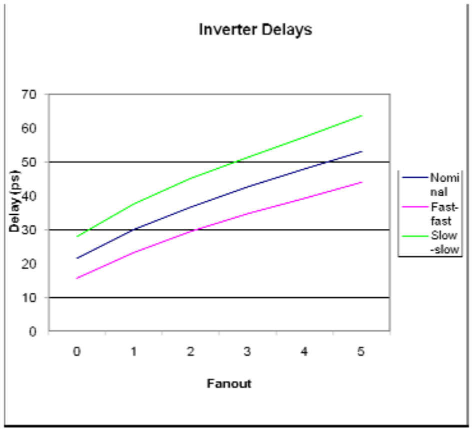
I. Sutherland, *et al*, Logical Effort, Academic Press, 1999

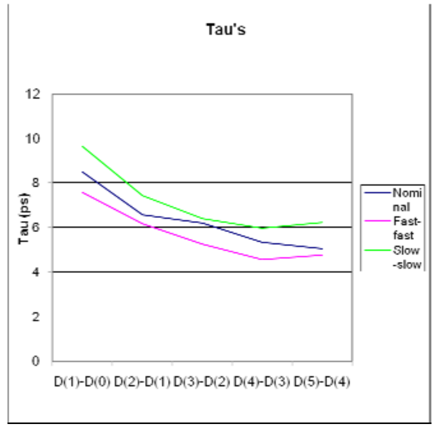
Wrong gate size



I. Sutherland, *et al*, Logical Effort, Academic Press, 1999

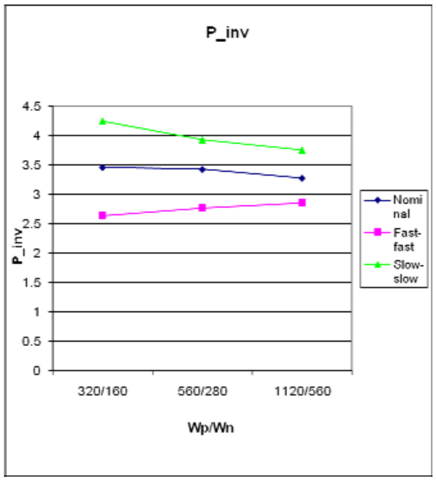
τ for our 130nm technology





I. Sutherland, *et al*, Logical Effort, Academic Press, 1999

Tau and Pinv for 130nm Tech



	Wp/Wn	Nom	Fast	slow
Average Tau value	320/160	6.26	5.74	6.98
	560/280	6.31	5.66	7.14
	1120/560	6.37	5.55	7.14
Average Pinv	320/160	3.46	2.64	4.24
	560/280	3.27	2.86	3.75
	1120/560	3.42	2.77	3.93

I. Sutherland, *et al*, Logical Effort, Academic Press, 1999

Best number of stages $p_{inv} = 3.38\tau$

Path effort F	Best number of stages $\hat{f}$	Min. delay $\hat{D}$	Stage effort f
0		3.38	
	1		0-9.57
9.57		12.9	
	2		3.09-7.38
54.4		21.5	
	3		3.79-6.65
294		30.1	
	4		4.14-6.29
1563		38.7	
	5		4.35-6.07
8246		47.3	
	6		4.49-5.93
43327		55.8	
	7		

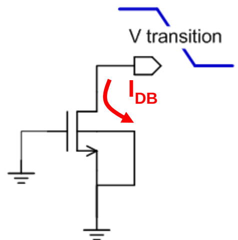
I. Sutherland, *et al*, Logical Effort, Academic Press, 1999

IBM 0.13um Characteristics

Char / Unit distance	Parameter
NMOS C_j	1.06fF/um (W)
PMOS C_j	1.03fF/um (W)
NMOS C_g	1.57fF/um (W)
PMOS C_g	1.44fF/um (W)
NMOS R_{eq}	43.24 kΩ/um (W)
PMOS R_{eq}	86.95 kΩ/um (W)

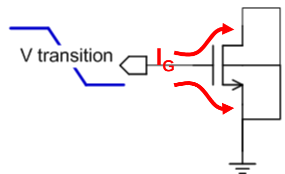
* All values obtained from device simulation

Device Simulations



Junction Capacitance

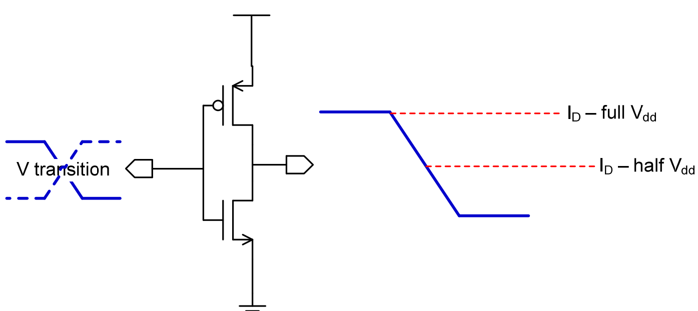
$$\int I\,dt = Q = C_j V$$



Gate Capacitance

$$\int I\,dt = Q = C_g V$$

Device Simulations



$$R_{eq} = \frac{R_{eq}^{full} + R_{eq}^{half}}{2}$$

$$R_{eq}^{full} = \frac{V_{DD}}{I_D^{full}} \qquad R_{eq}^{half} = \frac{V_{DD}}{I_D^{half}}$$

IBM 0.13um Characteristics

Char / Unit distance	Parameter	
Poly Resistance	58.33Ω/um (ℓ)	
M1 Resistance	0.44Ω/um (ℓ)	
M2-M6 Resistance	0.32Ω/um (ℓ)	
Upper layer Metals	0.09Ω/um (ℓ)	
	Plate Above, Below	**Isolated**
Poly Capacitance	0.1691 fF/um	0.1219 fF/um
M1 Capacitance	0.2619 fF/um	0.1565 fF/um
M2-M6 Capacitance	0.2341 fF/um	0.1592 fF/um
Upper layer Metals	0.1970 fF/um	0.1150 fF/um

* All values obtained from design manual

IBM 130nm Characteristics

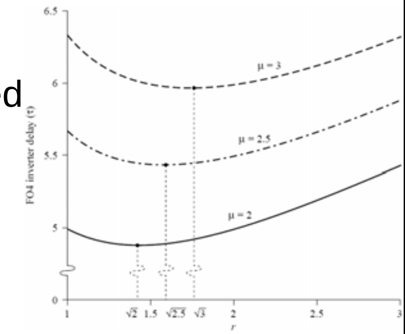
MOS Type	Parameter	Calculated	Simulated	% Error
NMOS	C_g	2.03fF/um	1.57fF/um	22.7%
PMOS	C_g	1.94fF/um	1.44fF/um	25.8%
NMOS	C_j	1.209fF/um	1.06fF/um	12.3%
PMOS	C_j	1.244fF/um	1.03fF/um	17.2%
NMOS	t_p (FO4)	42.8ps	35.5ps	21%
PMOS	t_p (FO4)	86.1ps	68.2ps	26%

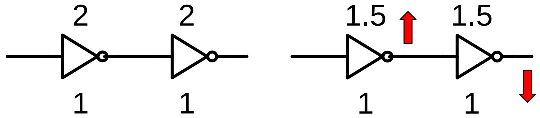
* Calculated $t_p = 0.69 * R_{eq} * C_L$
$C_L = 4C_{g,P} + 4C_{g,N} + C_{j,N} + C_{j,P}$

* Simulated t_p = 50% - 50% transition
FO4 delays for p/n ratio of 1 → should be adjusted

P/N ratio

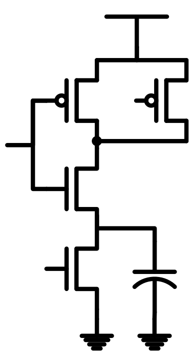
- Why use P/N = 2?
  - Noise margins are balanced
  - Equal slopes
- How about P/N = 1.5?

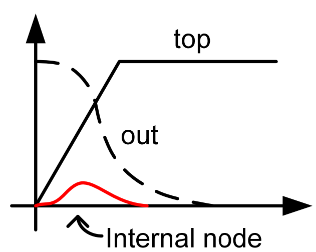




Limitations – Internal capacitance

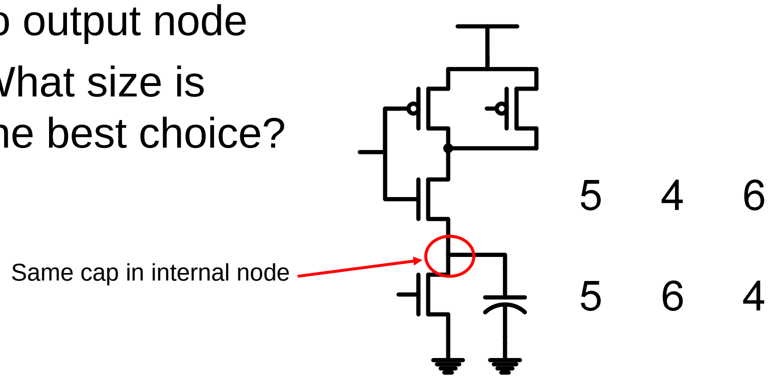
- Capacitance in internal nodes
- Body effect





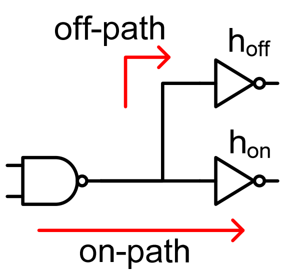
Limitations - Tapering

- Transistor sizes in stack are different
- The latest arriving input should be closest to output node
- What size is the best choice?



Limitations – Branching

- Assume that the size of the off-path gate tracks the size of the gate on-path



- Sizing one “critical” path of a branch may make the other paths worse

Limitations – Series Devices

- Models 2 series devices of size 2 to be equivalent in drive strength to a single device of size 1
  - Not really valid
- Velocity saturation
  - View 2 series devices as an equivalent device with 2W/2L
  - Devices with 2L have less velocity saturation effects, so current is greater W/L
- Solution – simulate gates to directly find logical efforts

Limitations – Slope & Interconnect

- Ignores impact of input slope on stage delay
- Interconnect capacitances hard to consider; requires iterations

Limitations - Scaling

- Scaling is not linear with width

$$d \approx RC$$

$$C = k_c S \Rightarrow C = k_c S + k_p$$ Perimeter

but really

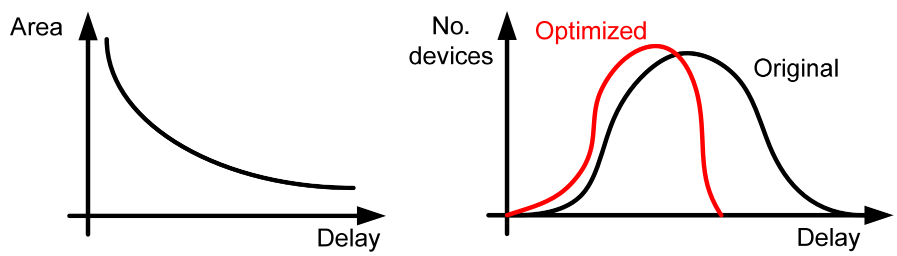
$$R = \frac{k_R}{S} \Rightarrow R = \frac{k_R}{S} + k_v$$ Narrow width effects
Process variations

- Parasitic delay of multi-input gates usually much less than simple model predicts
  - Diffusion sharing, input dependencies

Sizing tool

- Tool: TILOS [Dunlop 89]
  - Start with all transistors of min. size
  - Find critical path (Optimize path)
  - Compute delays
  - Increase size of transistors in “critical path”
  - Size path with best sensitivity
  - Repeat
  - Goal of path distribution → All paths equal in length…

Area - Delay



But the impact of process variations can be worse for the optimized paths – more on this later in the course

Summary

- Logical effort is useful for thinking of delay in circuits
  - NANDs are faster than NORs in CMOS
  - Paths are fastest when effort delays are ~4
  - Path delay is weakly sensitive to stages, sizes
  - But using fewer stages doesn't mean faster paths
  - Delay of path is about $\log_4 F$ FO4 inverter delays
  - Inverters and NAND2 best for driving large caps
- Provides language for discussing fast circuits
  - But requires practice to master