Introduction to CMOS VLSI Design

Chapter 4 Delay

thực tế trên mạch, về mặt tín hiệu, tín hiệu ngõ vào sẽ có tín hiệu ngõ ra tương ứng sau 1 khoảng thời gian nào đó nhất định, tức là có độ trễ

xét module logic hay module inverter có độ trễ rất nhỏ, vài chục ns, nhưng khi xem xét 1 con chip lớn có số lượng transistor tích hợp trên đó là cực kì lớn hoặc là con chip gồm nhiều module nhỏ ở trong, thì số lượng delay tích lũy cực kì lớn

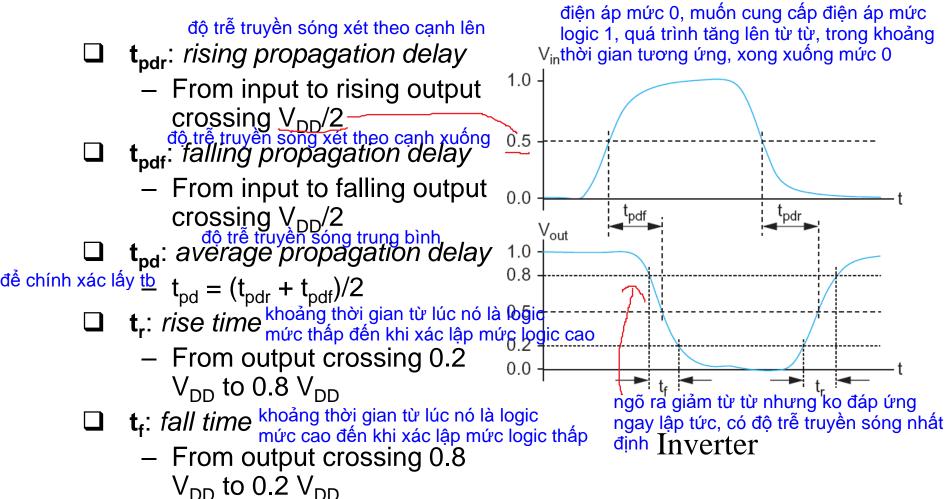
delay có nhiều dạng, ảnh hưởng nhiều nhất giữa ngõ vào và ngõ ra, ngta sẽ tính toán và gọi là propagation delay, thời gian delay này sẽ ảnh hưởng đến cái tần số hoạt động tối đa con chip của mình

f max: tần số hoạt động tối đa mà con chip có thể đáp ứng được

timing diagram: dạng sóng giản đồ thời gian thế hiện tín hiệu ngõ vào và tín hiệu ngõ ra, mối quan hệ như thế nào

Delay Definitions



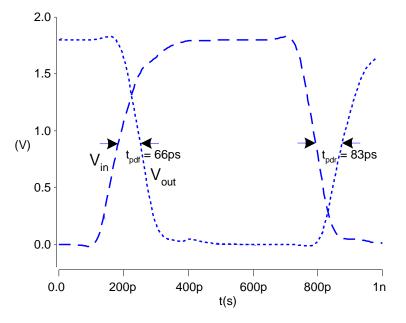


Delay Definitions

- □ t_{cdr}: rising contamination delay
 - From input to rising output crossing $V_{DD}/2$
- □ t_{cdf}: falling contamination delay
 - From input to falling output crossing V_{DD}/2
- □ t_{cd}: contamination delay
 - $-t_{cd} = (t_{cdr} + t_{cdf})/2$??
 - $t_{cd} = min(t_{cdr}, t_{cdf})$

Simulated Inverter Delay

- Solving differential equations by hand is too hard
- ☐ SPICE simulator solves the equations numerically
 - Uses more accurate I-V models too!
- ☐ But simulations take time to write, may hide insight



minh họa timing lệch như thế nào

Delay Estimation

- ☐ We would like to be able to easily estimate delay
 - Not as accurate as simulation
 - But easier to ask "What if?"
- ☐ The step response usually looks like a 1st order RC response with a decaying exponential.
- ☐ Use RC delay models to estimate delay
 - C = total capacitance on output node
 - Use effective resistance R
 - So that $t_{pd} = RC$
- ☐ Characterize transistors by finding their effective R
 - Depends on average current as gate switches

Effective Resistance

- ☐ Shockley models have limited value
 - Not accurate enough for modern transistors
 - Too complicated for much hand analysis
- ☐ Simplification: treat transistor as resistoroi cmos như những cái điện trở
 - Replace $I_{ds}(V_{ds}, V_{gs})$ with effective resistance R
 - $I_{ds} = V_{ds}/R$
 - R averaged across switching of digital gate
- ☐ Too inaccurate to predict current at any given time
 - But good enough to predict RC delay

RC Delay Model

dùng mạch tương đương để thể hiện transistor MOS

Use equivalent circuits for MOS transistors

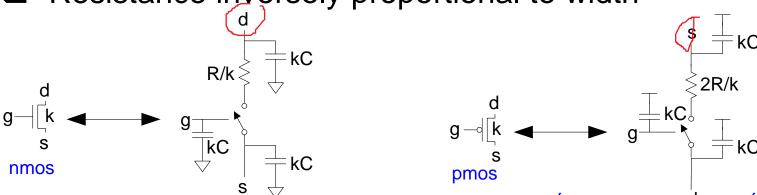
- switch lí tưởng, on hoặc off điện dung điện trở loeal switch + capacitance and ON resistance nếu 2 cái này đều xem xét ở kích thước giống nhau
 - Unit nMOS has resistance R, capacitance C độ linh động thấp hơn 2 lần, độ dẫn điện thấp hơn 2 lần, suy ra điện trở cao hơn 2 lần
 - Unit pMOS has resistance 2R (lower carrier slide 17 ch02

mobility), capacitance C

điện dung tương đương tỉ lệ thuận với kích thước transistor, kích thước tăng

Capacitance proportional to width (k) en k lần, điện dung tăng lên k lần điện trở tỉ lệ nghịch với kích thước, khi kích thước tăng lên, số hạt mang điện tăng lên, dòng điện cũng sẽ tăng lên

tương ứng luy Resistaince inversely proportional to width



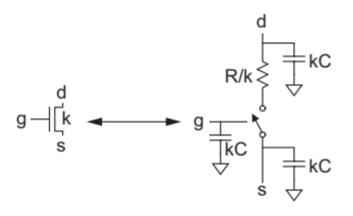
Chapter 4 CMOS VLSI Design 3 cái tụ kí sinh nối từ g xuống cái body, body nối GND

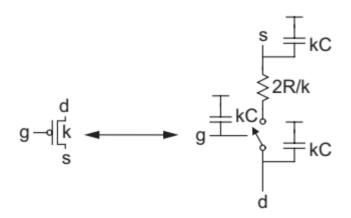
đế của transistor smos nối với Vdd, điện dung kí sinh tính từ cái cực đến Vdd

RC Values

- □ Capacitance
 - $-C = C_q = C_s = C_d = 2$ fF/ μ m of gate width in 0.6 μ m
 - Gradually decline to 1 fF/ μ m in nanometer techs. ((1 fF = 1.0E-15 F)
- Resistance
 - R ≈ 6 KΩ*μm in 0.6 μm process
 - Improves with shorter channel lengths
- Unit transistors
 - May refer to minimum contacted device (4/2 λ)
 - Or maybe 1 μm wide device
- Doesn't matter as long as you are consistent
 Chapter 4
 CMOS VLSI Design

Equivalent RC circuits



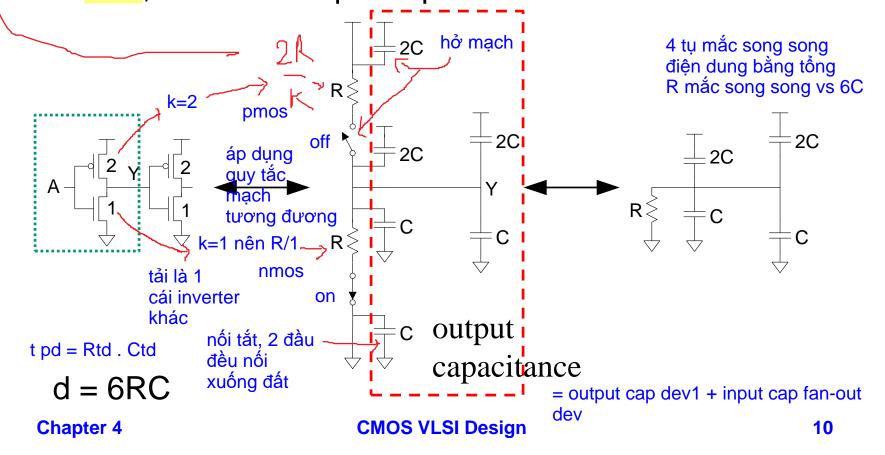


fanout 2 inverter and or ...

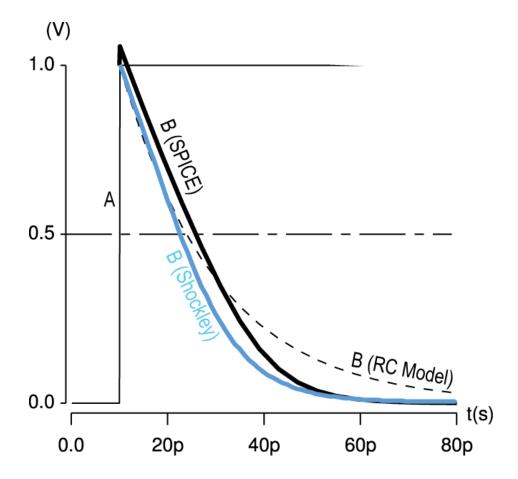
Inverter Delay Estimate

khi tính toán delay, delay phụ thuộc vào cái tải

- ☐ Estimate the delay of a fanout-of-1 inverter
- ☐ A=1, consider output capacitance



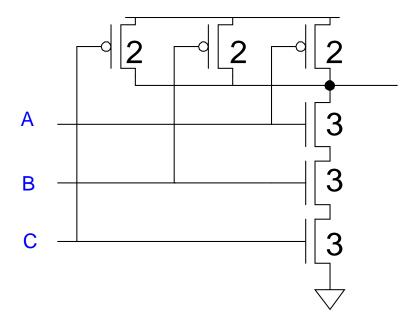
Delay Model Comparison (Example 4.1, p.145)



Example: 3-input NAND

phác họa

□ Sketch a 3-input NAND with transistor widths chosen to achieve effective rise and fall resistances equal to a unit inverter (R).



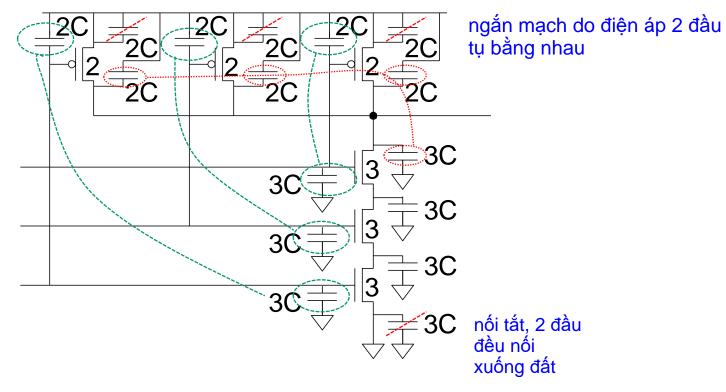
3-input NAND Caps

chú thích

Annotate the 3-input NAND gate with gate and diffusion capacitance.

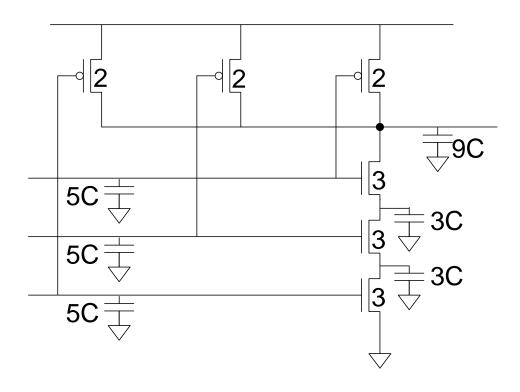
Các tụ điện mắc song song thì có điện dung tương đương bằng

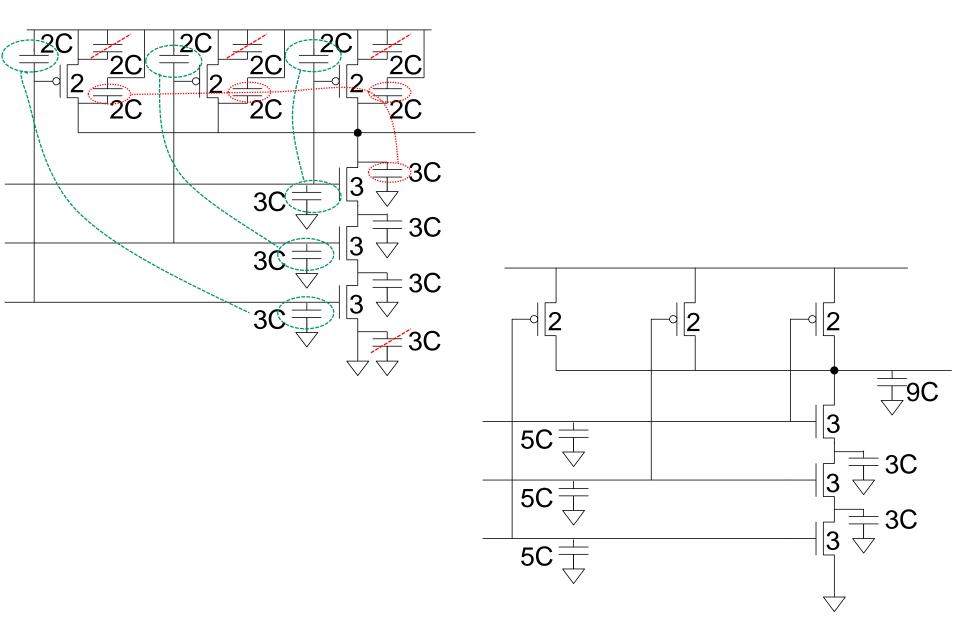
tổng điện dung của các tụ cộng lại



3-input NAND Caps

Annotate the 3-input NAND gate with gate and diffusion capacitance.





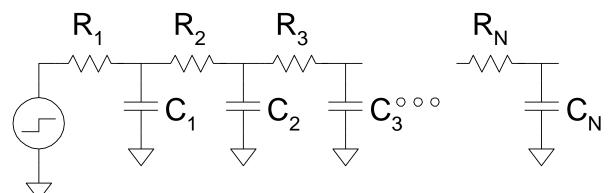
Elmore Delay

tu điện đầu ra

- ON transistors look like resistors (output capacitor) kéo lên kéo xuống
- ☐ Pullup or pulldown network modeled as RC ladder
- □ Elmore delay of RC ladder

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

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

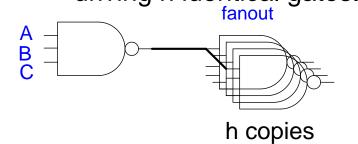


khi tính toán điện dung để ra cái điện dung tương đương, ngta bỏ đi Cg, do Y ngõ ra nằm ở cực g của mạch kế tiếp, do đó xem xét Cg ở tầng fanout

Example: 3-input NAND

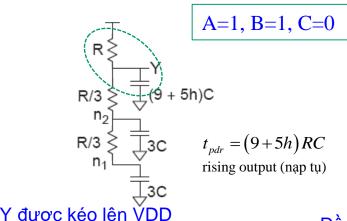
ngõ ra NAND 3-input nối với h cổng giống nhau

■ Estimate worst-case rising and falling delay of 3-input NAND driving h identical gates.



A 3 + 9C + 5hCB $3 + n_2 + 3C$ input cap of 1 fanout x h

bởi vì đầu ra chỉ được nạp điện qua R



Đầu ra kéo xuống thông qua ba transistor nMOS nối tiếp

Chapter 4

thông qua một transistor pMOS

CMOS VLSI Design

17

Delay Components

- Delay has two parts ky sinh
- ky sinh
 Parasitic delay
 15
 12
 9 or 11
 RC

 ky sinh là thời gian để một cổng điều khiển điện dung khuếch tán bên trong của chính nó. Tăng chiều rộng của transistor làm giảm điện trở nhưng tăng điện dung nên delay ký sinh là độc lập với kích thước cổng

 - Independent of load
 - delay nỗ lực phụ thuộc vào tỷ lệ h của điện dung tải bên ngoài đến điện dung - Effort delay đầu vào và do đó thay đổi theo độ rộng của transistor. Nó cũng phụ thuộc vào độ phức tạp của cổng
 - 5h RC

Tỷ lệ với điện dung tải

Proportional to load capacitance

Contamination Delay

delay ô nhiễm là delay nhanh nhất mà cổng có thể chuyển

- Best-case (contamination) delay (minimum delay) can be substantially less than propagation delay.
- Ex: If all three inputs fall simultaneously (rising output)

A
B
C
A=0, B=0, C=0

REPRESE

Y
$$t_{cdr} = \left[(9+5h)C \right] \left(\frac{R}{3} \right) = \left(3+\frac{5}{3}h \right)RC$$

$$\frac{1}{8} = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} \Rightarrow \text{Res}$$

$$\frac{1}{8} = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8} \Rightarrow \text{Res}$$

$$\frac{1}{8} = \frac{1}{8} \Rightarrow \text{Res}$$

$$\frac{1}{8} \Rightarrow \text{Res$$

rising output (nap tu)

Đối với quá trình chuyển đổi gia tăng, trường hợp tốt nhất là cả ba transistor pMOS đều bật đồng thời. Các transistor nMOS tắt, vì vậy n1 và n2 không được kết nối với ngõ ra và không góp phần làm delay

Chapter 4

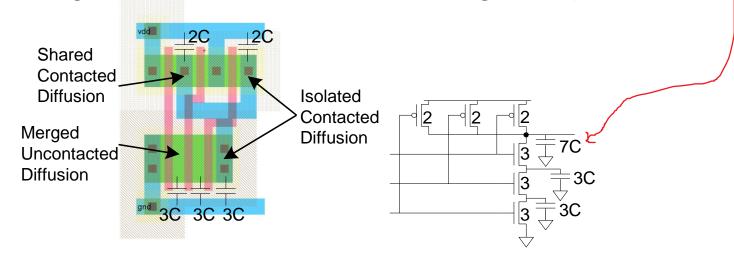
CMOS VLSI Design

19

Diffusion Capacitance

source/ drain

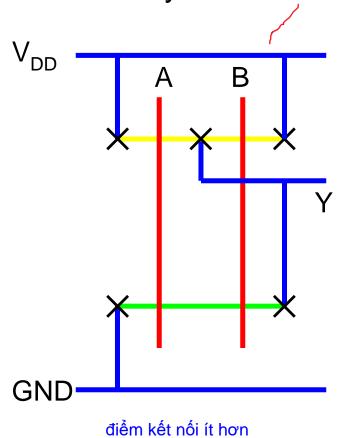
- We assumed contacted diffusion on every s / d.
- ☐ Good layout minimizes diffusion area
- □ Ex: NAND3 layout shares one diffusion contact
 - Reduces output capacitance by 2C giảm điện dung ngỗ ra 2C
 - Merged uncontacted diffusion might help too

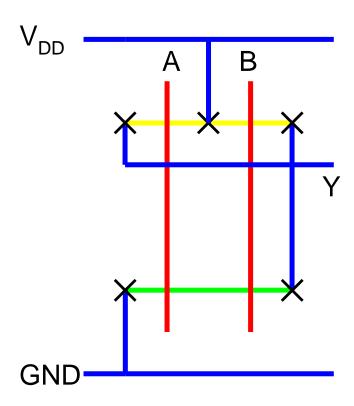


trang 153

Layout Comparison

Which layout is better?





Logical Effort Review

- □ Logical Effort
- □ Delay in a Logic Gate
- Multistage Logic Networks
- Choosing the Best Number of Stages
- Example
- □ Summary

Phép tính back-of-the-envelope là một phép tính thô, thường được ghi nhanh trên bất kỳ mẩu giấy vụn nào có sẵn như phong bì. Nó hơn một phỏng đoán nhưng ít chuẩn hơn một phép tính chính xác hoặc bằng chứng toán học. Đặc điểm xác định của các phép tính back-of-the-envelope bao gồm việc sử dụng các giả định đã được đơn giản hóa

Introduction

một loạt các lựa chọn hoang mang

- ☐ Chip designers face a bewildering array of choices
 - What is the best circuit topology for a function?
 - How many stages of logic give least delay?
 - How wide should the transistors be?
- Logical effort is a method to make these decisions
 - Uses a simple model of delay
 - Allows back-of-the-envelope calculations

lựa chọn thay thê

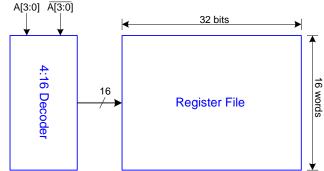
- Helps make rapid comparisons between alternatives
- Emphasizes remarkable symmetries

sự đối xứng

Example

□ A memory designer for an embedded automotive processor.
 Help design the decoder for a register file.

- Decoder specifications:
 - 16 word register file
 - Each word is 32 bits wide
 - Each bit presents load of 3 unit-sized transistors
 - True and complementary address inputs A[3:0]
 - Each input may drive 10 unit-sized transistors
- needs to decide:
 - How many stages to use?
 - How large should each gate be?
 - How fast can decoder operate?

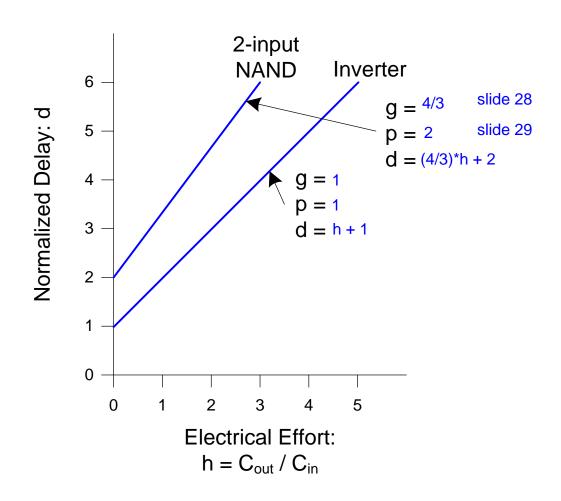


Delay in a Logic Gate

độc lập với quy trình Express delays in process-independent unit Delay has two components: d = f + ptau 3RC $\tau =$ f: effort delay = gh (a.k.a. stage effort) 3 ps in 65 nm process Again has two components 60 ps in 0.6 μm process g: logical effort Do khả năng cung cấp dòng điện tương đối của cổng
 Measures relative ability of gate to deliver current $-g \equiv 1$ for inverter h: electrical effort (or fanout) = C_{out} / C_{in} Ratio of output to input capacitance Sometimes called fanout p: parasitic delay (normally ~1) truyền động Represents delay of gate driving no load Set by internal parasitic capacitance

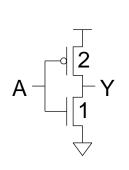
Delay Plots

$$d = f + p$$
$$= gh + p$$



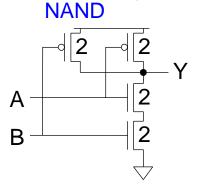
Computing Logical Effort

- □ DEF: Logical effort (g) is the ratio of the input capacitance of a gate to the input capacitance of an inverter delivering the same output current.
- Measure from delay vs. fanout plots
- Or estimate by counting transistor widths

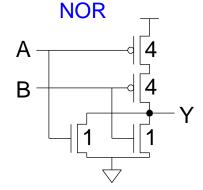


$$C_{in} = 3$$

 $g = 3/3$



$$C_{in} = 4$$
$$g = 4/3$$



$$C_{in} = 5$$

 $g = 5/3$

Catalog of Gates

g

☐ Logical effort of common gates

Gate type	Number of inputs					
	1	2	3	4	n	
Inverter	1					
NAND		4/3	5/3	6/3	(n+2)/3	
NOR		5/3	7/3	9/3	(2n+1)/3	
Tristate / mux	2	2	2	2	2	
XOR, XNOR		4, 4	6, 12, 6	8, 16, 16, 8		

Catalog of Gates

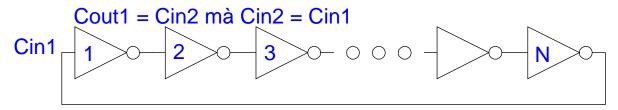
p

- ☐ Parasitic delay of common gates
 - In multiples of p_{inv} (≈1)

Gate type	Number of inputs					
	1	2	3	4	n	
Inverter	1					
NAND		2	3	4	n	
NOR		2	3	4	n	
Tristate / mux	2	4	6	8	2n	
XOR, XNOR		4	6	8		

Example: Ring Oscillator

Estimate the frequency of an N-stage ring oscillator



Logical Effort: g = 1

Electrical Effort: h = Cout / Cin = 1 vì nó truyền động một tải giống hệt nhau

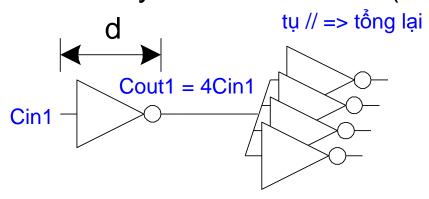
Parasitic Delay: p = 1

Stage Delay: d = gh + p = 1*1+1 = 2

Frequency: $f_{osc} = \frac{1}{(2*N*d)} = \frac{1}{(2*2*N)} = \frac{1}{(4N)}$

Example: FO4 Inverter

☐ Estimate the delay of a fanout-of-4 (FO4) inverter



Logical Effort: g = 1

Electrical Effort: $h = \frac{\text{Cout1}}{\text{Cin1}} = 4$

Parasitic Delay: p = 1

Stage Delay: d = gh + p = 1*4 + 1 = 5

Multistage Logic Networks

khái quát thành

nhiều tầng

- Logical effort generalizes to multistage networks
- Path Logical Effort

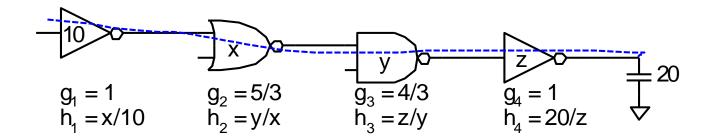
$$G = \prod g_i$$

□ Path Electrical Effort $H = \frac{C_{\text{out-path}}}{C_{\text{out-path}}}$

$$H = \frac{C_{\text{out-path}}}{C_{\text{in-path}}}$$

Path Effort

$$F = \prod f_i = \prod g_i h_i$$



Multistage Logic Networks

- ☐ Logical effort generalizes to multistage networks
- □ Path Logical Effort $G = \prod g_i$
- □ Path Electrical Effort $H = \frac{C_{out-path}}{C}$
- $lacksquare Path Effort <math>F = \prod f_i = \prod g_i h_i$
- Can we write F = GH?
 vấn đáp

Paths that Branch

□ No! Consider paths that branch:

= 2GH = 2*18 = 36

$$G = 1*1=1$$
 $H = 90/5 = 18$
 $GH = 18$
 $h_1 = (15+15)/5 = 6$
 $h_2 = 90/15 = 6$
 $C \text{ in path}$
 $C \text{ on path}$

Branching Effort

- ☐ Introduce *branching effort*
 - Accounts for branching between stages in path

$$b = \frac{C_{\text{on path}} + C_{\text{off path}}}{C_{\text{on path}}}$$

$$B = \prod b_i$$

Note:

$$\prod h_i = BH$$

■ Now we compute the path effort

$$-F = GBH$$

Multistage Delays

□ Path Effort Delay

$$D_F = \sum f_i$$

□ Path Parasitic Delay

$$P = \sum p_i$$

Path Delay

$$D = \sum d_i = D_F + P$$

Designing Fast Circuits

$$D = \sum d_i = D_F + P$$

Delay is smallest when each stage bears same effort

$$\hat{f} = g_i h_i = F^{\frac{1}{N}} \quad \sqrt{F}$$

Thus minimum delay of N stage path is

- This is a key result of logical effort
 - Find fastest possible delay
 - Doesn't require calculating gate sizes

Gate Sizes (~width k)

☐ How wide should the gates be for least delay?

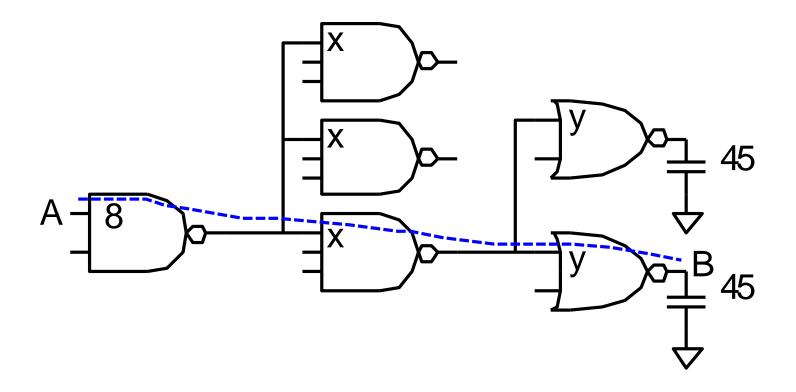
$$\hat{f} = gh = g \frac{C_{out}}{C_{in}}$$

$$\Rightarrow C_{in_i} = \frac{g_i C_{out_i}}{\hat{f}}$$
 kích thước cổng

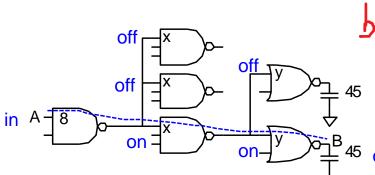
- □ Working backward, apply capacitance transformation to find input capacitance of each gate given load it drives.
- ☐ Check work by verifying input cap spec is met.

Example: 3-stage path

☐ Select gate sizes x and y for least delay from A to B

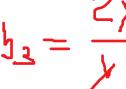


Example: 3-stage path



b = 8





tổng các C lại chia cho chính nó

Logical Effort

Electrical Effort

Branching Effort

Path Effort

Best Stage Effort

Parasitic Delay

Delay

G = tích g i =
$$(4/3) \times (5/3) \times (5/3) = 100/27$$
 dò catalog

$$H = C \text{ out-path } / C \text{ in-path} = 45/8$$

$$B = tich b i = 1*3*2 = 6$$

$$\hat{f} = F^{(1/N)} = 125^{(1/3)} = 5$$

$$P = sigma p i = 2 + 3 + 2 = 7$$
 dò catalog

D = Df + P = N *
$$F^{(1/N)}$$
 + P = 3*5 + 7 = 22 = 4.4 FO4

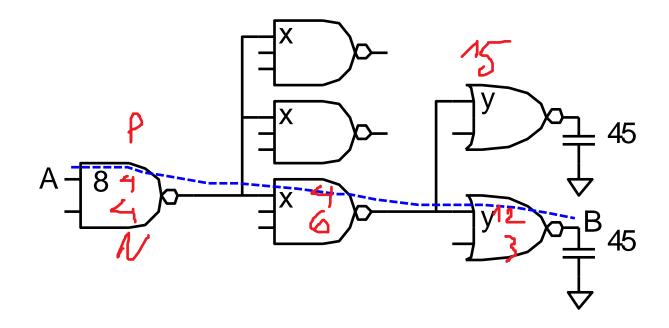
3 tầng

Example: 3-stage path

Work backward for sizes

$$y = gi * Cout i / f^ = (5/3) * 45 / 5 = 15$$

 $X = gi * Cout i / f^ = (5/3) * (15 + 15) / 5 = 10$



Best Number of Stages

- ☐ How many stages should a path use?
 - Minimizing number of stages is not always fastest
- ☐ Example: drive 64-bit datapath with unit inverter

```
do cùng 1 loại cổng inverter nên P = N \frac{1}{100} \frac
```

- Consider adding inverters to end of path
- $D = NF^{\frac{1}{N}} + \sum_{i=1}^{n_1} p_i + (N n_1) p_{inv}$ N-n₁ ExtraInverters $D = NF^{\frac{1}{N}} + \sum_{i=1}^{n_1} p_i + (N n_1) p_{inv}$ N-n₂ ExtraInverters $Path \ \text{Effort} F$

$$\frac{\partial D}{\partial N} = -F^{\frac{1}{N}} \ln F^{\frac{1}{N}} + F^{\frac{1}{N}} + p_{inv} = 0$$

lacksquare Define best stage effort $ho = F^{rac{1}{N}}$

$$p_{inv} + \rho (1 - \ln \rho) = 0$$

Best Stage Effort

 $p_{inv} + \rho (1 - \ln \rho) = 0 \text{ has no closed-form solution }$

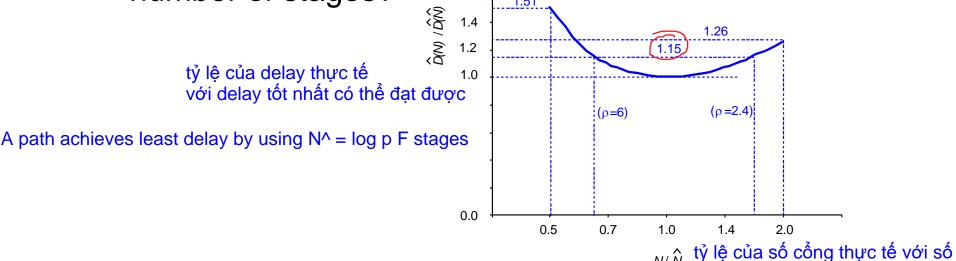
bỏ qua

- \square Neglecting parasitics (p_{inv} = 0), we find ρ = 2.718 (e)
- \Box For $p_{inv} = 1$, solve numerically for $\rho = 3.59$

Sensitivity Analysis

→ How sensitive is delay to using exactly the best number of stages?

→ 1.6 → 1.51



- \square 2.4 < ρ < 6 gives delay within 15% of optimal to 15% optimal to 15%
 - We can be sloppy!
 - Harris uses $\rho = 4$, exact ρ is process dependent

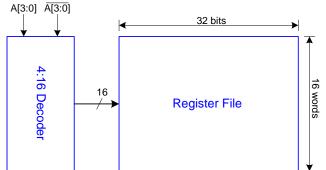
Optimal Power?

☐ Switching vs. Crowbar

Example, Revisited

□ A memory designer for an embedded automotive processor.
 Help design the decoder for a register file.

- Decoder specifications:
 - 16 word register file
 - Each word is 32 bits wide
 - Each bit presents load of 3 unit-sized transistors
 - True and complementary address inputs A[3:0]
 - Each input may drive 10 unit-sized transistors
- needs to decide:
 - How many stages to use?
 - How large should each gate be?
 - How fast can decoder operate?



Number of Stages

Decoder effort is mainly electrical and branching

Electrical Effort: $H = \frac{96}{10} = \frac{9.6}{10}$

Branching Effort: $B = \frac{(10 + 10^{*7})/10 = 8}{10^{*7}}$

☐ If we neglect logical effort (assume G = 1)

Path Effort: F = GBH = 1*8*9.6 = 76.8

Number of Stages: $N = \log p F = \log 476.8 = 3$

□ Try a -stage design (But G≠1 and ρ≠4)

— word [0]

Gate Sizes & Delay

G =
$$1*(6/3)*1 = 2$$
 H = $96/10 = 9.6$
B = $1*8*1 = 8$

$$H = 96/10 = 9.6$$
 $R = 1*8*1 = 8$

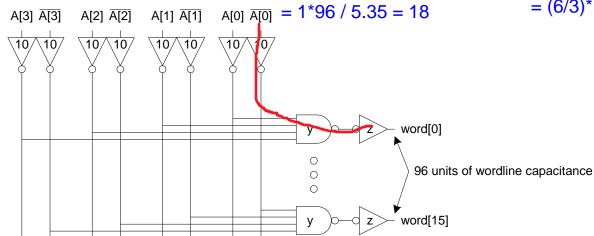
$$f = F^{(1/N)} = 153.6^{(1/3)} = 5.35$$

$$N = N * F^{(1/N)} + P = 3*5.35 + (1+4+1) = 22.06$$

$$Z = g i * Cout i / f^$$

$$y = gi * Cout i / f^{\land}$$

= (6/3)*18 / 5.35 = 6.7



1 cổng A nối với 8 chân cổng NAND

Comparison

bảng tính

☐ Compare many alternatives with a spreadsheet

 \Box D = N(76.8 G)^{1/N} + P

Design	N	G	Р	D
NOR4	1	3	4	234
NAND4-INV	2	2	5	29.8
NAND2-NOR2	2	20/9	4	30.1
INV-NAND4-INV	3	2	6	22.1
NAND4-INV-INV	4	2	7	21.1
NAND2-NOR2-INV-INV	4	20/9	6	20.5
NAND2-INV-NAND2-INV	4	16/9	6	19.7
INV-NAND2-INV-NAND2-INV	5	16/9	7	20.4
NAND2-INV-NAND2-INV-INV	6	16/9	8	21.6

vấn đáp

Review of Definitions

Term	Stage	Path
number of stages	1	N
logical effort	g	$G = \prod g_i$
electrical effort	$h = \frac{C_{ m out}}{C_{ m in}}$	$H = rac{C_{ ext{out-path}}}{C_{ ext{in-path}}}$
branching effort	$b = \frac{C_{\text{on-path}} + C_{\text{off-path}}}{C_{\text{on-path}}}$	$B = \prod b_i$
effort	f = gh	F = GBH
effort delay	f	$D_F = \sum f_i$
parasitic delay	p	$P = \sum p_i$
delay	d = f + p	$D = \sum d_i = D_F + P$

Method of Logical Effort

1) Compute path effort

F = GBH

2) Estimate best number of stages

 $N = \log_4 F$

- 3) Sketch path with N stages
- 4) Estimate least delay

 $D = NF^{\frac{1}{N}} + P$

5) Determine best stage effort

 $\hat{f} = F^{rac{1}{N}}$

6) Find gate sizes

 $C_{in_i} = \frac{g_i C_{out_i}}{\hat{f}}$

Limits of Logical Effort

- ☐ Chicken and egg problem
 - Need path to compute G
 - But don't know number of stages without G
- □ Simplistic delay model
 - Neglects input rise time effects
- ☐ Interconnect
 - Iteration required in designs with wire
- Maximum speed only
 - Not minimum area/power for constrained delay

Summary

- ☐ Logical effort is useful for thinking of delay in circuits
 - Numeric logical effort characterizes gates
 - 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 log₄F FO4 inverter delays
 - Inverters and NAND2 best for driving large caps
- □ Provides language for discussing fast circuits
 - But requires practice to master