ECE4740: Digital VLSI Design

Lecture 6: Inverter sizing

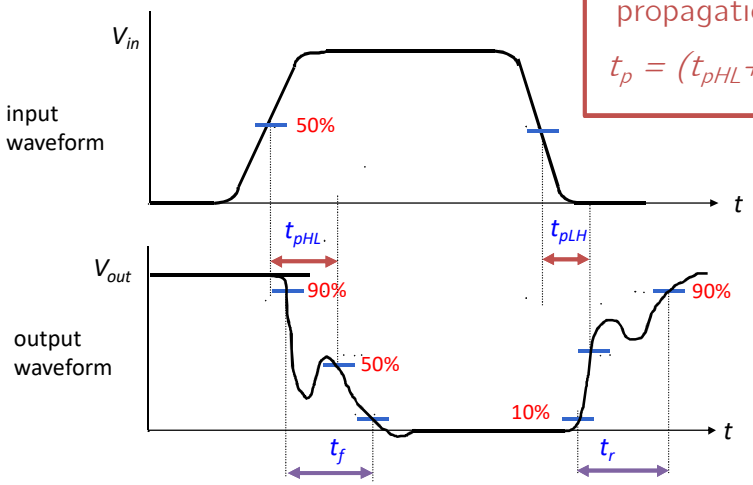
193

Very important concepts

Brief timing recap

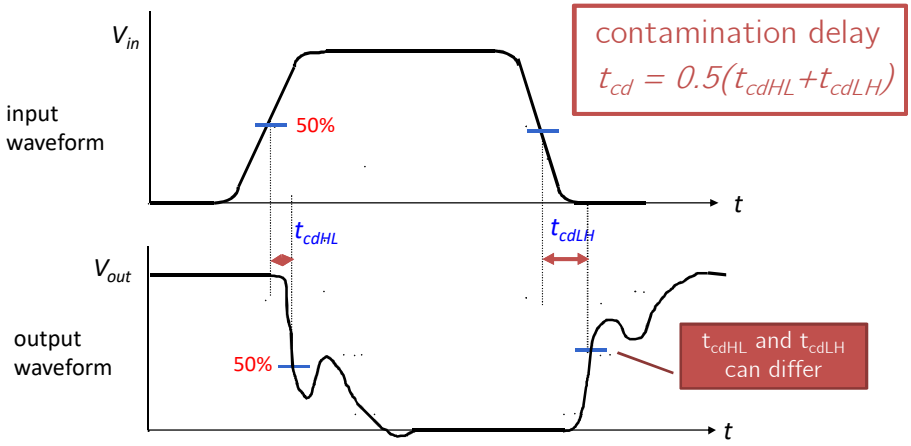
194

Rise/fall times & propagation delay



195

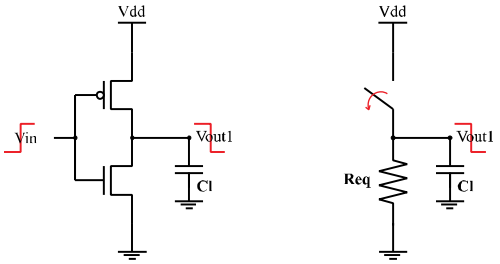
Contamination delay



196

First-order analysis: extract R_{eq}

- Approximations required as $C_L(v)$ and i(v) are non-linear and voltage v dependent



$$R_{eq} \approx \frac{3}{4}\frac{V_{DD}}{I_{DSATn}}\left(1-\frac{5}{6}\lambda V_{DD}\right)$$

$$I_{DSATn} = \mu_n C_{ox}\frac{W_n}{L_n}\left((V_{DD}-V_{Tn})V_{DSATn}-\frac{V_{DSATn}^2}{2}\right)$$

197

Use RC circuit to model t_{pHL}

- Decay time from 100% to 50% is

$$t_{pHL} = \ln(2)R_{eqn}C_L \approx 0.69R_{eqn}C_L$$

- Same can be obtained for low-to-high time

$$t_{pLH} = \ln(2)R_{eqp}C_L \approx 0.69R_{eqp}C_L$$

- Propagation delay:

$$t_p = \frac{t_{pHL}+t_{pLH}}{2} \approx 0.69C_L\left(\frac{R_{eqn}+R_{eqp}}{2}\right)$$

198

Increase both W_n and W_p!

Inverter sizing

199

Split load capacitance C_L

$$t_p \approx 0.69 R_{eq}(C_{int} + C_{ext})$$

intrinsic output capacitances from self-loading etc. (diffusion, Miller)

extrinsic capacitances: fan-out and wiring

- Just rewrite as:

$$t_p \approx 0.69 R_{eq} C_{int} \left(1 + \frac{C_{ext}}{C_{int}}\right)$$

t_{p0}: "unloaded" propagation delay

effect on propagation delay caused by extrinsic capacitances

assume that NMOS and PMOS are sized such that rise and fall times are equivalent

200

Impact of sizing factor S

$$t_p \approx 0.69 R_{eq} C_{int} \left(1 + \frac{C_{ext}}{C_{int}}\right)$$

- Intrinsic capacitances: $C_{int} = SC_{ref}$
- Gate resistance: $R_{eq} = R_{ref}/S$

capacitance of a unit-sized device

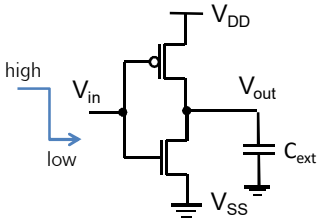
resistance of a unit-sized device

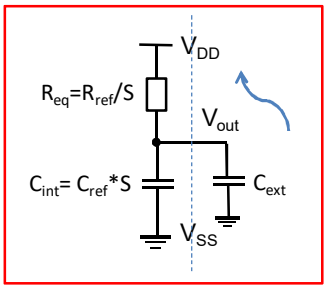
- Leads to:

$$t_p \approx 0.69 R_{ref} C_{ref} \left(1 + \frac{C_{ext}}{SC_{ref}}\right)$$

201

An equivalent view





- Propagation delay: $t_p = 0.69 R_{eq}(C_{int} + C_{ext})$
- Simplifying:

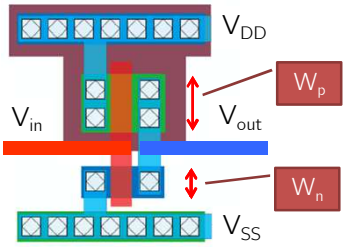
$$t_p = 0.69 \frac{R_{ref}}{S}(C_{ref}S + C_{ext})$$

$$= 0.69 R_{ref} C_{ref} \left(1 + \frac{C_{ext}}{SC_{ref}}\right)$$

202

What exactly is "S"?

- The sizing factor is how much larger you make the widths of NMOS and PMOS
  - You usually keep the PMOS:NMOS ratio fixed
  - Gate lengths L usually kept at the minimum



S=2 corresponds to 2Wp and 2Wn

203

Impact of sizing S (cont'd)

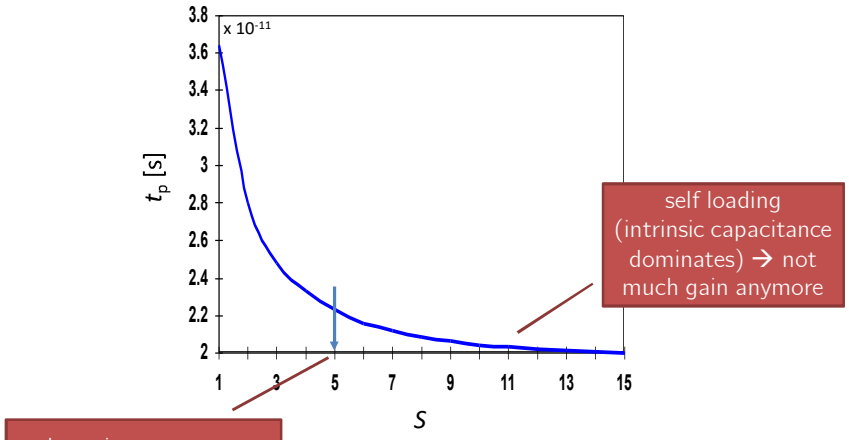
$$t_p \approx 0.69 R_{ref} C_{ref} \left(1 + \frac{C_{ext}}{S C_{ref}}\right)$$

intrinsic delay of inverter: t_{p0}

- Intrinsic delay of gate independent of S
  - no load means no effect on propagation delay
- Making S large, eliminates effect of load
  - at the cost of inverter area!
  - and it will affect the load of the preceding logic!

204

Example: Sizing impact* on t_p



self loading (intrinsic capacitance dominates) → not much gain anymore

large improvement already achieved for S=5

Image taken from:
http://bwrcs.eecs.berkeley.edu/Classes/icdesign/ee141_f01/Notes/chapter5.pdf

*for a fixed load!

205

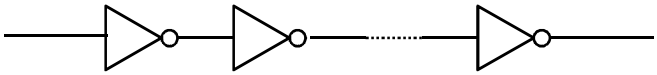
A more relevant case

Sizing a chain of inverters

206

Sizing a chain of inverters

- For isolated inverter: Increasing S reduces delay but also increases input capacitance
  - not very useful in practice!
- More relevant case: chain of inverters:



207

It's a trade-off

- Sizing up an inverter reduces delay, but will also increase its input capacitance!
- Intrinsic capacitance C_{int} proportional to gate capacitance: $C_{int} = \gamma C_g$

$$t_p \approx \underbrace{0.69 R_{eq} C_{int}}_{=t_{p0}} \left(1 + \frac{C_{ext}}{C_{int}}\right)$$

- γ is technology dependent (and about 1)

208

The effective fan-out f

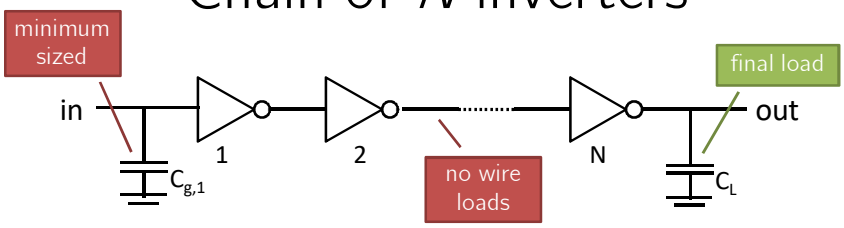
- Delay of an inverter: a function of the ratio between external load and its input cap!

$$t_p \approx t_{p0}\left(1 + \frac{C_{ext}}{\gamma C_g}\right) = t_{p0}(1 + f/\gamma)$$

- Effective fan out: $f = \frac{C_{ext}}{C_g}$
- The goal is to minimize the delay through the entire inverter chain

209

Chain of N inverters



- Delay of j[th] inverter: $t_{p,j} = t_{p0}\left(1 + \frac{C_{g,j+1}}{\gamma C_{g,j}}\right)$
- Total delay is: $t_p = \sum_{j=1}^{N} t_{p,j}$
- Assume: $C_{g,N+1} = C_L$
- Optimality conditions: $\delta t_p / \delta C_{g,j} = 0$

210

Image adapted from: Digital Integrated Circuits (2nd Edition) by Rabaey, Chandrakasan, Nikolic

Optimum size of inverter

$$\frac{\delta}{\delta C_{g,j}}\left(1+\frac{C_{g,j}}{\gamma C_{g,j-1}}\right)+\left(1+\frac{C_{g,j+1}}{\gamma C_{g,j}}\right)=0$$

- Optimum size of each inverter is geometric mean of neighbor's sizes!

$$\frac{C_{g,j}}{C_{g,j-1}}=\frac{C_{g,j+1}}{C_{g,j}} \qquad C_{g,j}=\sqrt{C_{g,j-1}C_{g,j+1}}$$

211

Implications

$$\frac{C_{g,j}}{C_{g,j-1}}=\frac{C_{g,j+1}}{C_{g,j}}=f_{opt} \Rightarrow C_{g,j+1}=f_{opt}C_{g,j}$$

- Each inverter should be sized up by same factor f_{opt} compared to preceding gate

$$C_L=f_{opt}^N C_{g,1} \qquad \text{WHY?}$$

212

Implications (cont'd)

- With $f_{opt} = \sqrt[N]{C_L/C_{g,1}} = \sqrt[N]{F}$

  the minimum (total) delay is

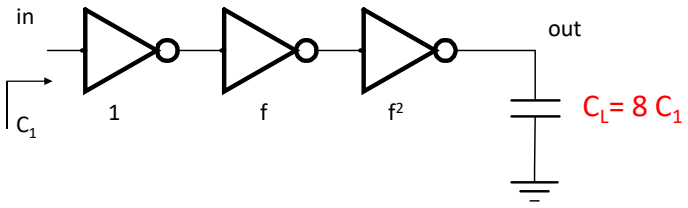
$$t_{p,min} = N t_{p0}(1 + \sqrt[N]{F}/\gamma)$$

- Simply a result of

$$t_p = \sum_{j=1}^{N} t_{p,j} \qquad t_{p,j} = t_{p0}(1 + f_{opt}/\gamma)$$

213

Example: simple chain



- F=C_L/C_1 has to be evenly distributed across N=3 stages: $f_{opt} = \sqrt[3]{8} = 2$

$$t_{p,min} = 3t_{p0}(1 + \sqrt[3]{8}/\gamma)$$

214

Optimal number of stages N_{opt}

$$t_{p,min} = N t_{p0}(1 + \sqrt[N]{F}/\gamma)$$

- There is a trade-off
  - too many stages: intrinsic delay dominates
  - too few stages: effective fan-out dominates

$$\frac{\delta}{\delta N} N(\gamma + \sqrt[N]{F}) = 0$$

- Has no closed-form solution ☹
- Solution is about 3.6 for $\gamma = 1$

215

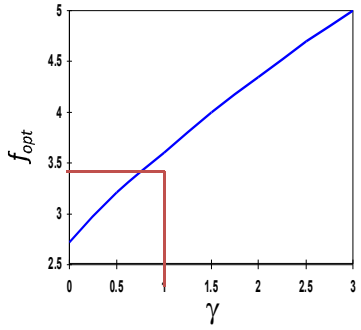
Assume no self loading

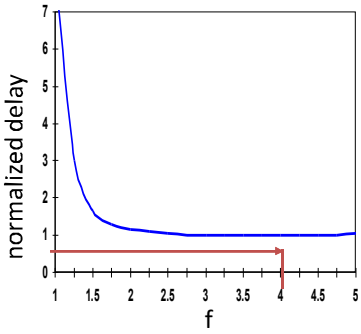
- Implies $\gamma = 0$
- Solve $\frac{\delta}{\delta N} \log\left(N \sqrt[N]{F}\right) = 0$
- Leads to so-called exponential horn

$$N_{opt} = \log(F) \quad \Rightarrow \quad f_{opt} = e \approx 2.72$$

216

Optimum effective fan-out f_{opt}



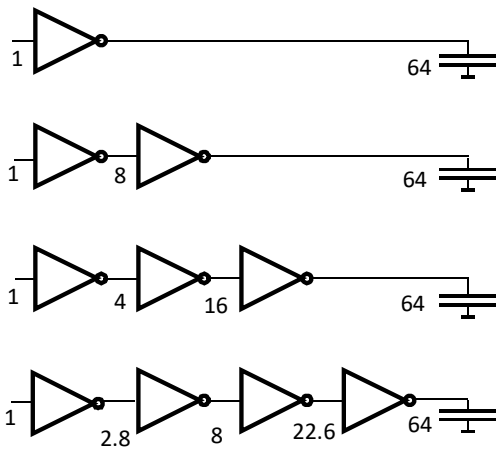


- Choosing f larger than optimum has little effect on delay and reduces # stages
  - f=4 is common practice for $\gamma = 1$
  - too many stages has negative impact on delay

217

Image taken from: http://bwrcs.eecs.berkeley.edu/Classes/icdesign/ee141_f01/Notes/chapter5.pdf

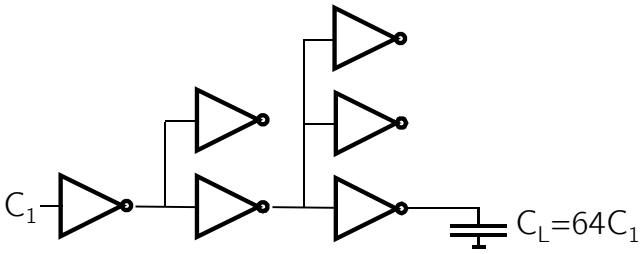
Example: buffer design



N	f	t_p
1	64	65
2	8	18
3	4	15
4	2.8	15.3

218

Example: size the buffers

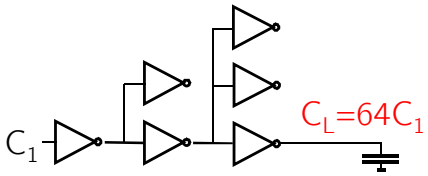


always C_{ext}/C_{int}

$$\frac{2C_2}{C_1} = \frac{3C_3}{C_2} = \frac{C_L}{C_3} = f_{opt}$$

$$\frac{2C_2}{C_1} \times \frac{3C_3}{C_2} \times \frac{C_L}{C_3} = f_{opt}^3 \quad \Rightarrow \quad \frac{6C_L}{C_1} = f_{opt}^3$$

Example: size the buffers (cont'd)



$$\sqrt[3]{\frac{6C_L}{C_1}} = f_{opt} \approx 7.2685$$

buffer 2 is 3.6x larger than buffer 1

$$\frac{2C_2}{C_1} \approx 7.2685 \Rightarrow C_2 \approx 3.6342C_1$$

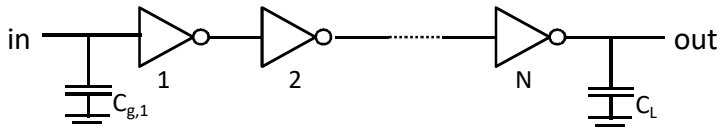
buffer 3 is 8.8x larger than buffer 1

$$\frac{3C_3}{C_2} \approx 7.2685 \Rightarrow C_3 \approx 2.4228C_2 \Rightarrow C_3 \approx 8.8051C_1$$

Examples

Delay minimization

Two cases for delay minimization



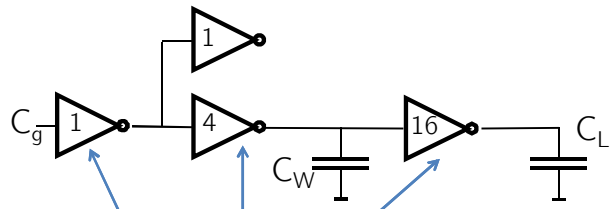
- Known number of gates N, unknown f
  - Find optimal f such that $F=f^N$
- Known f_{opt}, unknown N
  - Find N with f as close as possible to f_{opt}

$$f_{opt} = \sqrt[N]{C_L/C_{g,1}} = \sqrt[N]{F}$$

$$t_p = Nt_{p0}(1 + \sqrt[N]{F}/\gamma)$$

Image adapted from: Digital Integrated Circuits (2nd Edition) by Rabaey, Chandrakasan, Nikolic

Example: propagation delay t_p



$$t_p = t_{p1} + t_{p2} + t_{p3}$$

$$t_{p1} = t_{p0}\left(1 + \frac{4C_g + C_g}{\gamma C_g}\right) \qquad t_{p2} = t_{p0}\left(1 + \frac{16C_g + C_W}{\gamma 4C_g}\right)$$

$$t_{p3} = t_{p0}\left(1 + \frac{C_L}{\gamma 16C_g}\right)$$

223