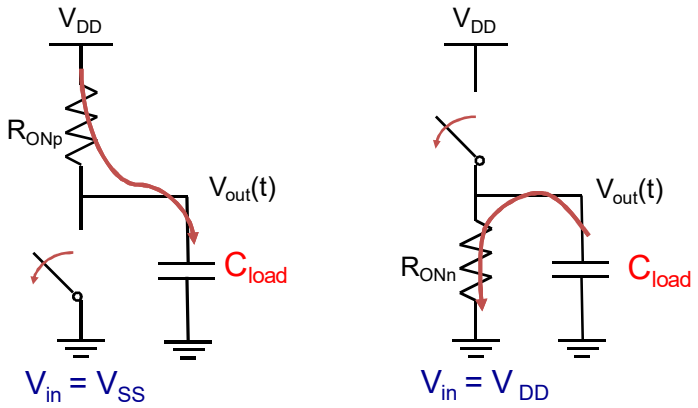
ECE4740: Digital VLSI Design

Lecture 5: Dynamic behavior of inverter

RC circuits ahead!

Dynamic behavior

Dynamic behavior: intuition

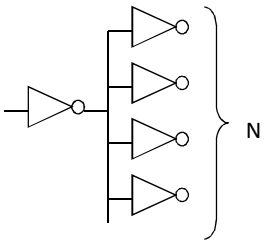


- $V_{out}(t)$ depends on $\tau = R_{on}C_{load}$

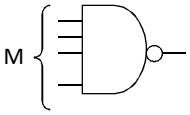
163

C_{load} depends on fan-in and fan-out

- Fan-out: number of load gates connected to output of driving gate
  - All wiring and total input capacitance determine C_{load}

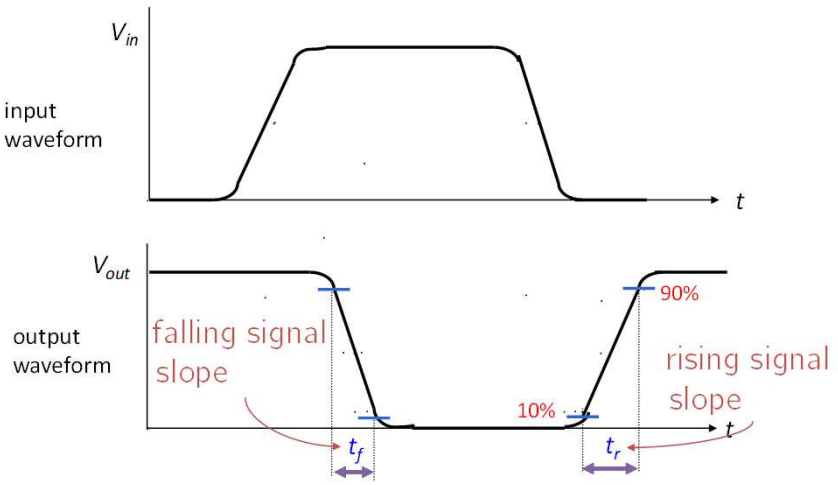


- Fan-in: number of inputs to gate



164

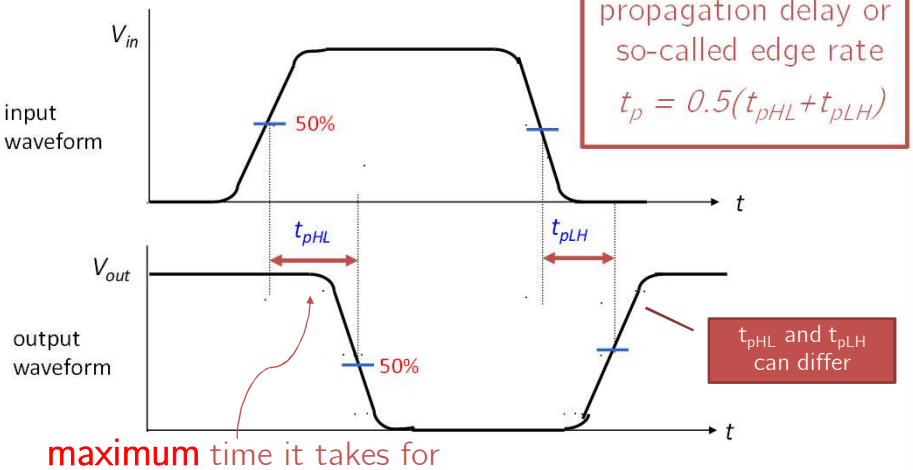
Rise and fall times*



*always measured at the output

165

Propagation delay



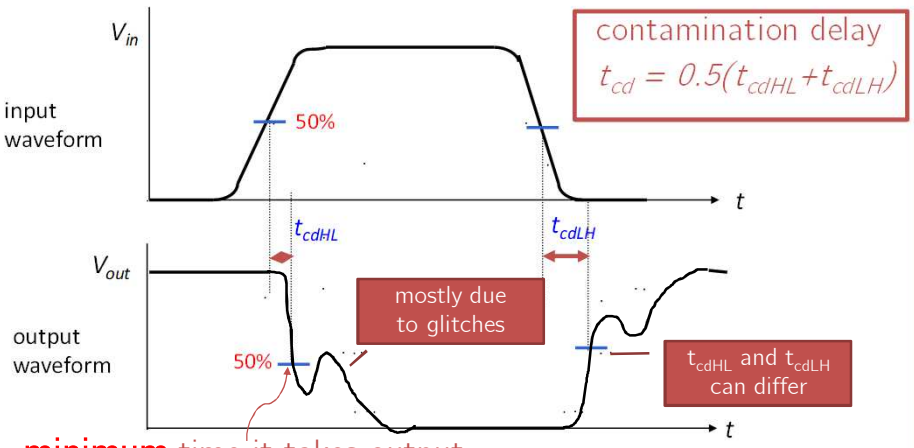
propagation delay or so-called edge rate

$t_p = 0.5(t_{pHL}+t_{pLH})$

maximum time it takes for output to cross 50% voltage

166

Contamination delay or min-delay



contamination delay
$t_{cd} = 0.5(t_{cdHL}+t_{cdLH})$

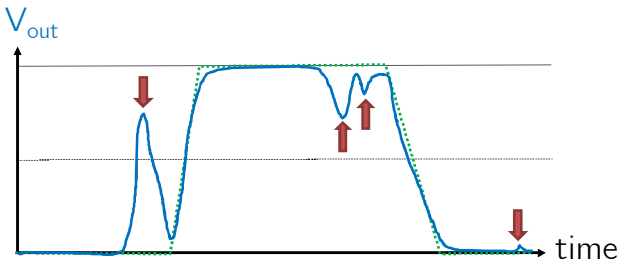
minimum time it takes output to cross 50% voltage

we have: $t_{cd}<=t_{pd}$

167

Electronic glitch

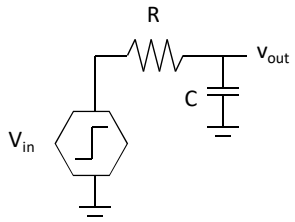
- Undesired transition that occurs before the signal settles to intended value
- Glitches can go from rail to rail
- Caused by logic (can be prevented)



168

RC model for dynamic behavior*

- First-order RC model
- Step function V_{SS} → V_{DD}



Rise time (10% to 90%) is

$$t_r \approx 2.2RC$$

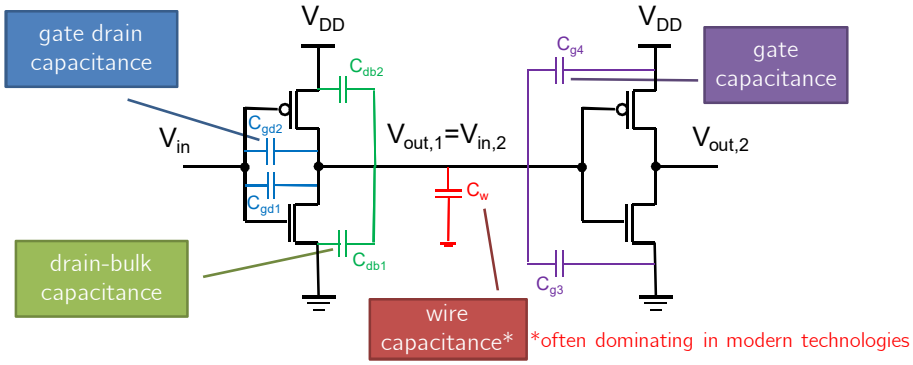
$$V_{out}(t) = V_{dd}(1 - e^{-t/\tau}) \qquad \tau = RC$$

*simple RC model is unable to model electronic glitches

169

Reality is more complicated

- Parasitic capacitances that affect transient behavior of cascaded inverter pair:

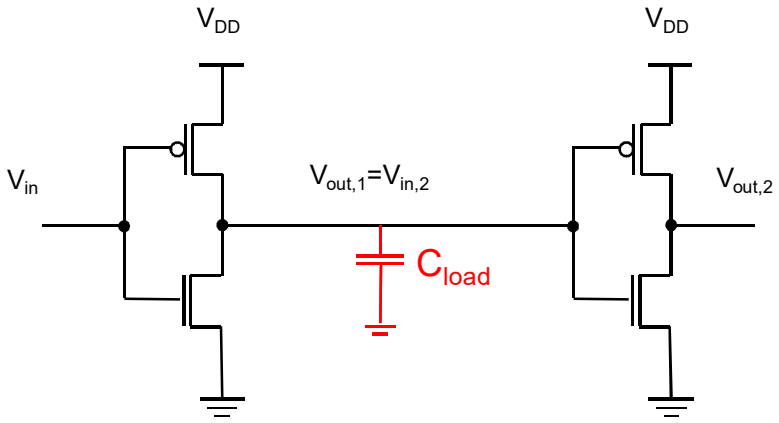


*often dominating in modern technologies

170

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

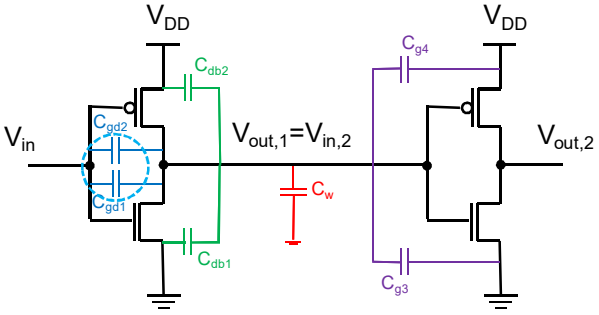
Idea: lump everything into C_{load}



171

Lump everything into C_{load}

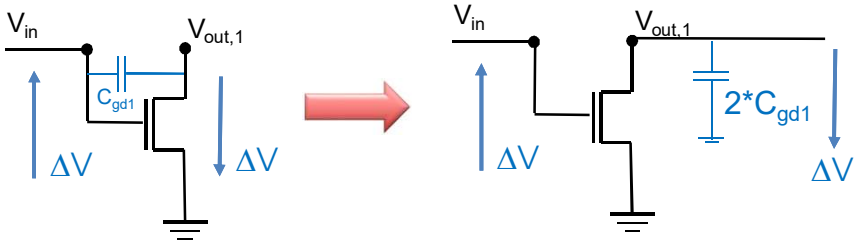
- Use the so-called Miller effect for $C_{gd1,2}$
- Linearize all non-linear capacitances



172

Miller capacitance*

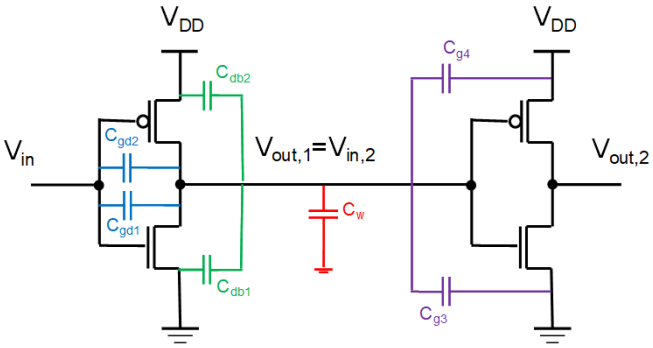
- Miller effect
  - Capacitor experiencing identical but opposite voltage swings at both its terminals
  - Replace with cap to GND of twice the C



*Miller effect depends on gain: $C_M=(1-g)*C_{in}$ (we assume g=-1)

173

Lumped load capacitance



$$C_{load} = 2C_{gd1} + 2C_{gd2} + C_{db1} + C_{db2} + C_{g3} + C_{g4} + C_w$$

- Only consider capacitances that switch!

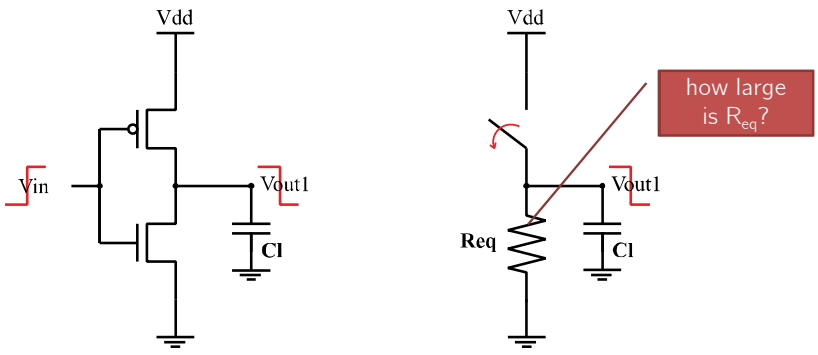
174

Will determine maximum clock frequency

Propagation delay

First-order analysis

- Exact computation difficult as $C_L(v)$ and $i(v)$ are non-linear in voltage $v(t)$



Computing the average on resistance

- Integrate over time:

average between t_1 and t_2

$$R_{\text{eq}} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} R_{\text{on}}(t)\text{d}t$$

- Resistance is equal to: $R_{\text{on}}(t) = \frac{V_{DS}(t)}{I_{DS}(t)}$

- But we are lazy and approximate it as:

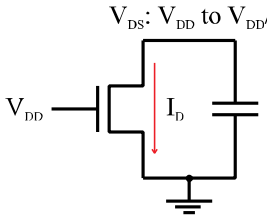
$$R_{\text{eq}} \approx \frac{1}{2}\left(R_{\text{on}}(t_1) + R_{\text{on}}(t_2)\right)$$

average at endpoints t_1 and t_2

177

Simple approximation of R_{eq}

- Discharge of capacitor from V_{DD} to GND
- Propagation delay: 50% discharge = $V_{DD}/2$
- Idea: average resistance at endpoints



$$R_{eq} = \frac{1}{2}\left(\frac{V_{DD}}{I_{DSATn}(1+\lambda V_{DD})} + \frac{V_{DD}/2}{I_{DSATn}(1+\lambda V_{DD}/2)}\right)$$

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

178

Intuition on R_{eq}

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

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

- Increasing width W reduces R_{eq}
- Increasing length L increases R_{eq}
- If V_{DD} → V_{Tn} then R_{eq} increases a lot!
- For $V_{DD} \gg V_{Tn} + V_{DSAT}/2$ resistance is almost independent of V_{DD}

179

Use RC circuit to model t_{pHL}/t_{pLH}

- Decay time from 100% to 50% is

any ideas why?

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

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

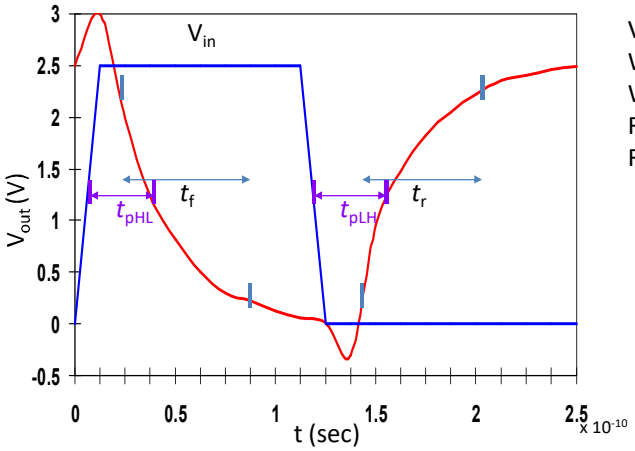
$$t_{pLH} = \ln(2) R_{eqp} C_L \approx 0.69 R_{eqp} C_L$$

- Propagation delay:

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

180

Real inverter transient response



V_{DD}=2.5V
W/L_n = 1.5
W/L_p = 4.5
R_{eqn}= 13kΩ
R_{eqp}= 31kΩ

t_{pHL} = 36ps

t_{pLH} = 29ps

so

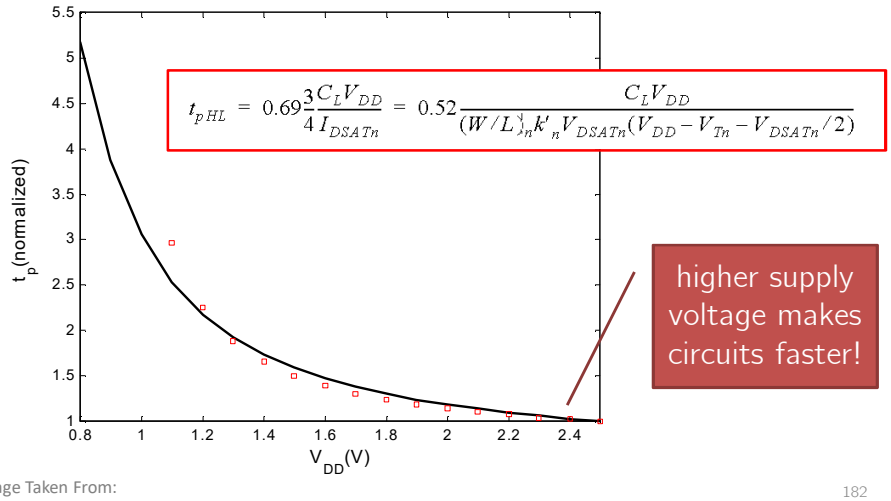
t_p = 32.5ps

From simulation: t_{pHL} = 39.9ps and t_{pLH} = 31.7ps for 0.25μm process

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

181

Propagation delay t_p vs. V_{DD}



$$t_{pHL} = 0.69\frac{3}{4}\frac{C_L V_{DD}}{I_{DSATn}} = 0.52\frac{C_L V_{DD}}{(W/L)_n k'_n V_{DSATn}(V_{DD} - V_{Tn} - V_{DSATn}/2)}$$

higher supply voltage makes circuits faster!

Image Taken From:
http://bwrcs.eecs.berkeley.edu/Classes/icdesign/ee141_f01/Notes/chapter5.pdf

182

Example: Intel's speed step tech.

- Adjust voltage and clock frequency to operation conditions
- Used in virtually all processors nowadays
- Why would someone do that?

Image taken from: http://news.softpedia.com/news/Intel-Geneseo-vs-AMD-Torrenza-36730.shtml

183

Ways to reduce t_{pHL}

- Increase V_{DD}: trade energy vs. performance
- Assume

$$V_{DD} \gg V_{Tn} + V_{DSAT}/2$$

$$t_{pHL} \approx 0.52 \frac{C_L}{(W_n/L_n)\mu_n C_{ox} V_{DSATn}}$$

- Reduce C_L: fanout, interconnect, C_{DS}
- Increase W/L ratio: careful with self loading

184

Another way to reduce t_{pHL}

- Carrier mobility μ_n depends on temperature
- Again, assume $V_{DD} \gg V_{Tn} + V_{DSAT}/2$

$$t_{pHL} \approx 0.52 \frac{C_L}{(W_n/L_n)\mu_n C_{ox} V_{DSATn}}$$

- For electrons in Si, we have $\mu_n \propto T^{-2.4}$
- Increasing the temperature reduces carrier mobility*, which increases propagation delay

*due to increased scattering; for holes we have $\mu_p \propto T^{-2.2}$

185

Cooling makes CMOS faster

- Known to gaming enthusiasts who overclock their CPUs → liquid cooling

nominal clock is 4.3 GHz in turbo mode

CPU Frequency World Record ranking on 3 February 2018

RANK	SCORE	USER	PROCESSOR	COOLING
#1	8722.78 MHz	The Stilt	AMD FX-8370 @ 8722.8MHz	Liquid Nitrogen
#2	8709 mhz	AndreYang	AMD FX-8150 @ 8709MHz	Liquid Nitrogen
#3	8659.64 mhz	Smoke	AMD FX-8370 @ 8659.6MHz	Liquid Nitrogen
#4	8615.39 mhz	slamms	AMD FX-8350 @ 8615.4MHz	Liquid Nitrogen
#5	8543.71 mhz	wytiwx	Intel Celeron D 352 @ 8543.7MHz	Liquid Nitrogen

http://hwbot.org/benchmark/cpu_frequency/halloffame

186

We can make gates faster

Minimizing the propagation delay

187

PMOS/NMOS ratio

- One approach: Size PMOS and NMOS so that R_{eq}'s match (ratio of about 2-to-3)
  - leads to symmetric VTC
  - optimizes noise margins
  - leads to equal t_{pHL} and t_{pLH}
- How about minimizing t_p?

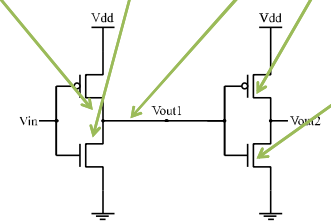
188

Minimizing t_p

- Widening PMOS
  - reduces R_{eqp}, which reduces t_{pLH}
  - also increases parasitic C_L, which increases t_{pHL}
- There is an optimal PMOS width!

$$C_L = (C_{dp1} + C_{dn1}) + C_W + (C_{gp2} + C_{gn2})$$

load capacitance of 1st gate



189

Minimizing t_p (cont'd)

$$C_L = (C_{dp1} + C_{dn1}) + C_W + (C_{gp2} + C_{gn2})$$

- Assume PMOS β times larger than NMOS

$$C_{dp1} \approx \beta C_{dn1} \qquad C_{gp2} \approx \beta C_{gn2} \qquad \beta = \frac{(W/L)_p}{(W/L)_n}$$

- Propagation delay becomes

$$t_p \approx \frac{0.69}{2}\left((1+\beta)(C_{dn1} + C_{gn2}) + C_W\right)\left(R_{eqn} + \frac{R_{eqp}}{\beta}\right)$$

190

After some math…

- Optimal value of β is

$$\beta_{opt} = \sqrt{\frac{R_{eqp}}{R_{eqn}}\left(1 + \frac{C_W}{C_{dn1} + C_{gn2}}\right)}$$

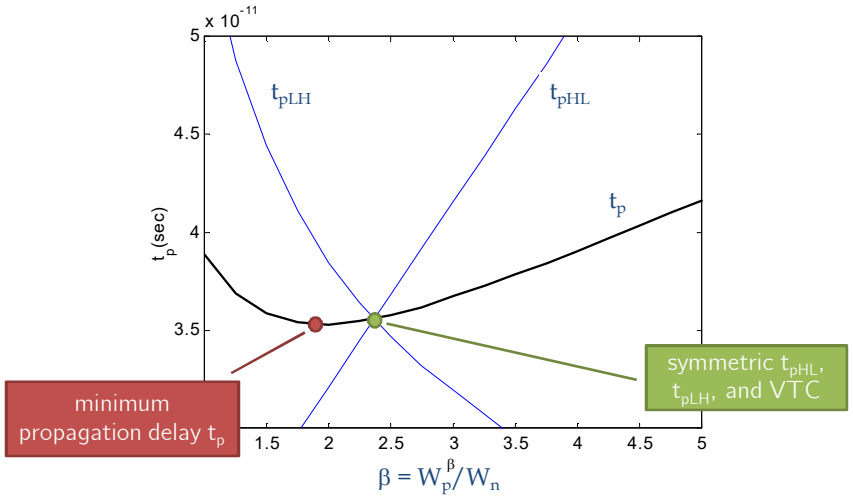
ratio *r* of symmetric VTC-sized resistors

- if C_W small, then you can ignore this
- If C_W large, then larger PMOS needed

- Interesting: smaller PMOS reduces t_p (at the cost of VTC symmetry and noise margin)

191

Example



192

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