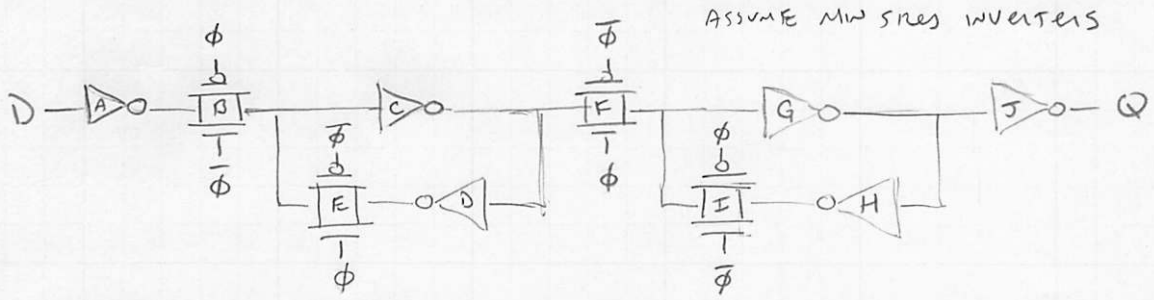
CMOS SEQUENTIAL STATE

DELAY
- SETUP
- PROPAGATION
- HOLD

ENERGY
- DATA VS. CLOCK
- CLOCK GATING

AREA → TOTAL GATE CAP

BASIC TRANSMISSION GATE MASTER/SLAVE FLIPFLOP



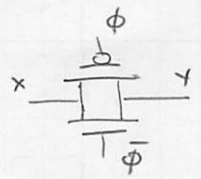
WE WANT TO CHARACTERIZE THE SETUP TIME, CLOCK-TO-Q PROPAGATION DELAY AND THE HOLD TIME

LOGICAL EFFORT WILL NOT DIRECTLY BE OF MUCH USE SINCE THE TRANSMISSION GATES COUPLE THE DELAY OF VARIOUS GATES TOGETHER

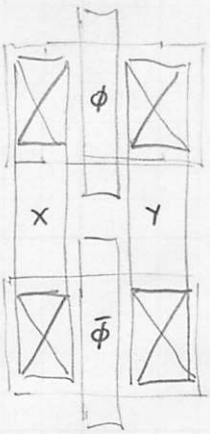
SO WE WILL USE ELMORE DELAY, BUT FIRST WE NEED AN RC MODEL OF A TRANSMISSION GATE

NOTE THE CONSERVATIVE USE OF EXTRA INVERTERS AT THE INPUT AND OUTPUT. THIS IS COMMON IN STANDARD CELLS TO HELP IMPROVE THE CELL'S "ELECTRICAL MODULARITY"

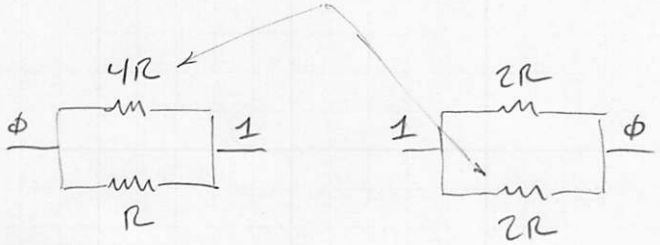
RC MODEL FOR TRANSMISSION GATE



ASSUME SIZE OF NMOS AND PMOS ARE THE SAME MINIMUM WIDTH



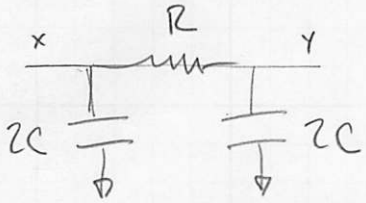
ASSUME A TRANSISTOR DRIVING A WEAK VALUE (IE PMOS PASSING ZERO, NMOS PASSING ONE) HAVE 2x WORSE EFFECTIVE RESISTANCE



$$R_{EFF} = \frac{R_1 R_2}{R_1 + R_2} = \frac{4R^2}{5R} = \frac{4}{5}R$$

$$R_{EFF} = \frac{R_1 R_2}{R_1 + R_2} = \frac{4R^2}{4R} = R$$

NOTE THAT INPUT NODE HAS PARASITIC CAP OF 2C AND OUTPUT NODE ALSO HAS PARASITIC CAP OF 2C



ROUGHLY ASSUME EFFECTIVE RESISTANCE IS R INDEPENDENT OF WHICH VALUE WE ARE PASSING

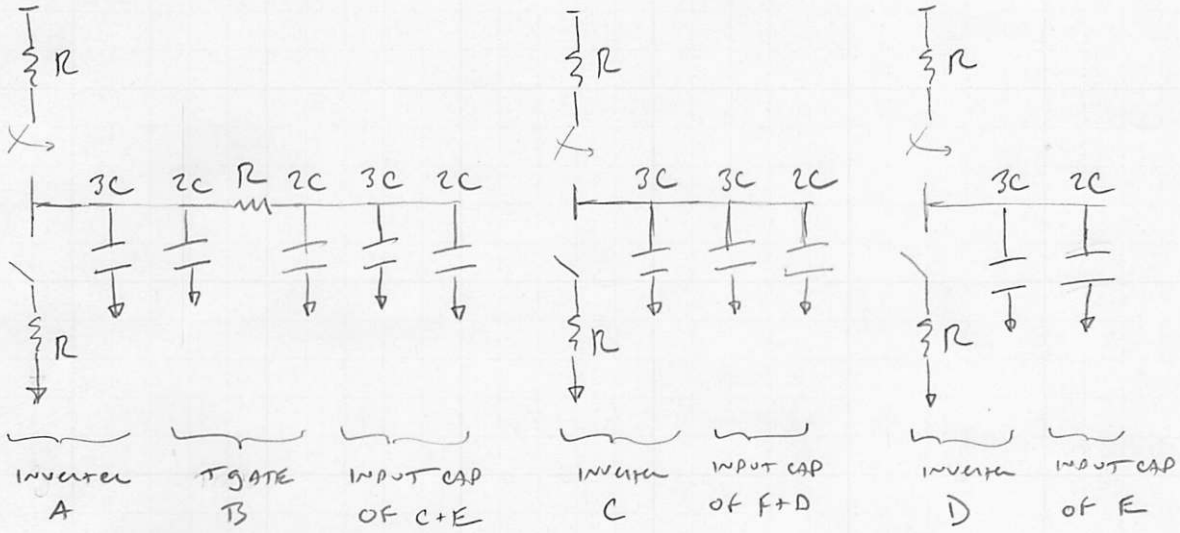
NOW THAT WE HAVE AN RC MODEL FOR A TRANSMISSION GATE WE CAN GO BACK AND CREATE RC MODELS FOR THE SETUP TIME, CLOCK-TO-Q PROPAGATION DELAY, AND HOLD TIME.

SETUP TIME

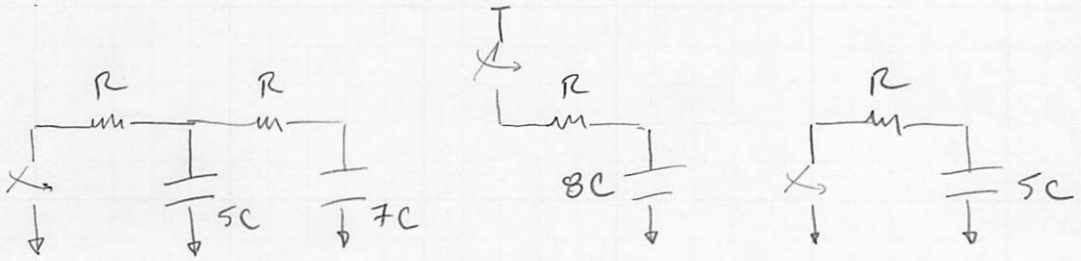
THE SETUP TIME IS THE LATEST THE INPUT CAN CHANGE WHILE STILL ALLOWING US TO CAPTURE THE VALUE.

TO CALCULATE SETUP TIME WE ASK OURSELVES, "HOW FAR DOES THE INPUT SIGNAL NEED TO PROPAGATE SO THAT WE CAN RELIABLY "FLIP" THE MASTER LATCH BEFORE THE CLOCK EDGE.

IF WE LOOK AT OUR SCHEMATIC WE CAN CONSERVATIVELY SAY THE INPUT SIGNAL MUST GO THROUGH GATES A, B, C, AND D TO REALLY FLIP THE CROSSCOUPLED INVERTERS.



ASSUME INPUT IS ONE, AND SIMPLIFY.

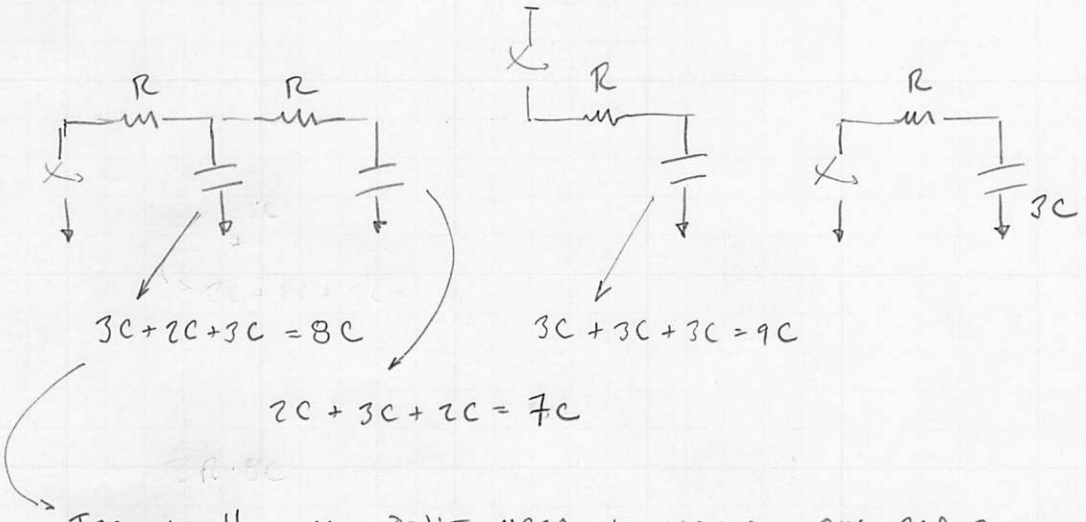


5RC + 14RC 8RC 5RC

SETUP TIME IS 32RC / 3RC = 9.7 T

Clock-to-Q propagation Delay

After the rising clock edge TGATE B and I open and TGATE E and F close. The propagation delay is from gate C through gates F, G, and J



Technically we don't need to include this cap since it was already discharged during the setup time

Clock to Q propagation Delay is

$$7R \cdot 7C + 9RC + 3RC = 26RC \,|\, 3RC = 8.6\tau$$

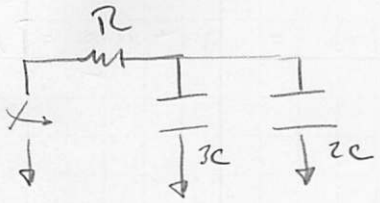
Hold Time

Hold time is how long we need to keep the input state after the sampling edge (eg. rising clock edge in this case) in order to prevent corrupting the state

If we assume ϕ and $\overline{\phi}$ change instantaneously on the rising clock edge then we actually have a negative hold time. If the input changes right after the edge then by the time it gets to the first transmission gate (gate B) that gate is already open and the input signal cannot corrupt the state.

In fact, the input can change a little before the edge since it takes some time to propagate through the first inverter.

More specifically, the delay through the first inverter is just



$$5RC = \frac{5RC}{3RC} = 1.67\,\tau$$

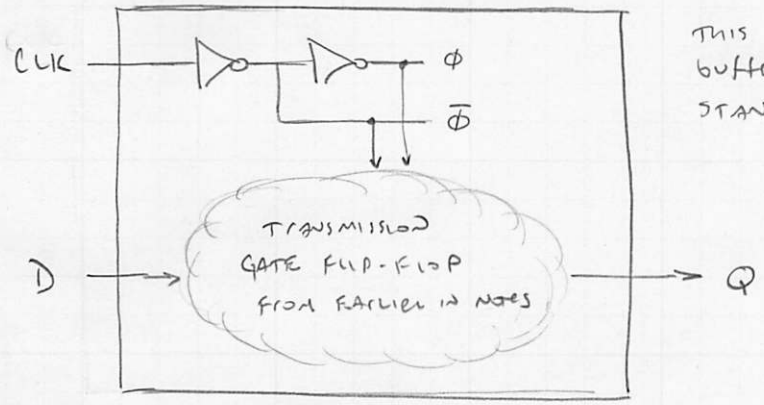
Of course we get the same answer using logical effort

$$d = gh + p = (1)\left(\frac{2C}{3C}\right) + 1 = 1.67\,\tau$$

So the hold time if we assume ϕ and $\overline{\phi}$ change instantly on the rising clock edge is $-1.67\,\tau$

Internal Clock Delay

What if we assume there is some delay between the clock input pin of our cell and the actual ϕ and $\overline{\phi}$ signals?

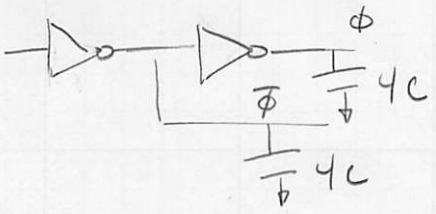


This kind of intra-cell clock buffering is not unusual in standard cells.

How does this impact the setup time, propagation delay, and hold time? Remember that all three of these metrics are defined with respect to the CLK pin of the cell NOT the local ϕ and $\overline{\phi}$ signals

First we need to calculate the delay through the local clock tree and then we can factor this into the setup time, propagation delay, and hold time

ϕ AND $\overline{\phi}$ EACH DRIVE ONE MINIMUM SIZED TRANSISTOR IN FOUR TRANSMISSION GATES, SO THE LOAD FOR EACH ONE IS 4C



WE CAN CONSERVATIVELY FOCUS ON JUST THE PATH TO ϕ SINCE IT WILL BE LONGER THAN THE PATH TO $\overline{\phi}$ AND DOMINATE WHEN THE TRANSMISSION GATES OPEN OR CLOSE

$$
\begin{aligned}
d &= g_0 h_0 + p_0 + g_1 h_1 + p_1 \\
&= (1)\left(\frac{7C}{3C}\right) + 1 + (1)\left(\frac{4C}{3C}\right) + 1 \\
&= \frac{7}{3} + 1 + \frac{4}{3} + 1 \\
&= \frac{17}{3} = 5.6\,\tau
\end{aligned}
$$

NOTICE WE SIMPLY INCORPORATE THE EXTRA FIXED CAP IN THE ELECTRICAL EFFORT. WE ARE NOT OPTIMIZING SIZING JUST CALCULATING DELAY

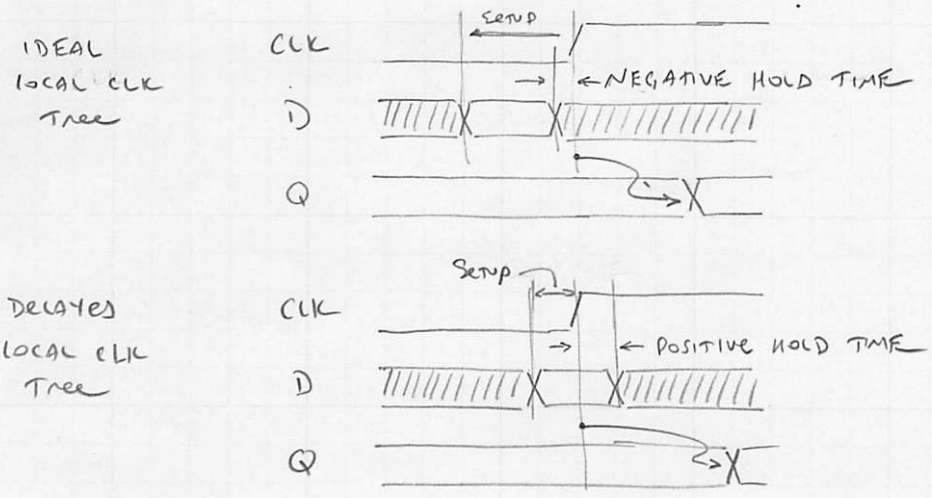
THIS EXTRA DELAY REDUCES THE SETUP TIME, INCREASES THE PROPAGATION DELAY, AND INCREASES THE HOLD TIME

SETUP TIME $9.7 - 5.6 = 4.1\,\tau$

CLK-TO-Q DELAY $8.6 + 5.6 = 14.2\,\tau$

HOLD TIME $-1.7 + 5.6 = 3.9\,\tau$ ← NOW WE HAVE A POSITIVE HOLD TIME

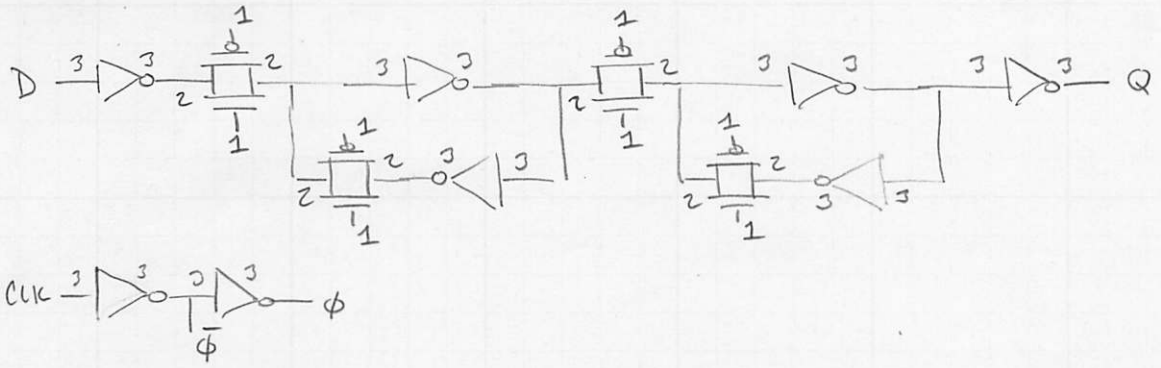
ESSENTIALLY WHAT WE HAVE DONE IS SHIFT THE SAMPLING WINDOW



ENERGY

We will look at the energy due to toggling the DATA lines vs the energy due to toggling the clock lines

First let's label all of the gate and parasitic caps



We are interested in the worst case avg energy per write/read access so we can estimate this by simply calculating the maximum switched capacitance while assuming every node toggles

$$\text{DATA switched cap} = 6\times6 + 4\times4 = 52C$$

$$\text{CLOCK switched cap} = 2\times6 + 4\times2 = 20C$$

Remember that for 90nm, $C \approx 0.5fF$, so energy can be estimated as follows

$$E_{\text{DATA TRANSITION}} = \frac{1}{2} C V_{DD}^2 = \frac{1}{2}(52C)\frac{0.5fF}{C}(1V)^2 = 13fJ$$

$$E_{\text{CLOCK TRANSITION}} = \frac{1}{2} C V_{DD}^2 = \frac{1}{2}(20C)\frac{0.5fF}{C}(1V)^2 = 5fJ$$

Which consumes more total energy? total power?

Depends on activity factor and clock frequency

LET'S ASSUME A CLOCK RATE OF 500 MHZ (NOT UNREASONABLE FOR 90 nm ASIC PROCESSOR) AND AN ACTIVITY FACTOR OF 0.1 FOR DATA AND 2 FOR CLOCK

NOTE THAT THE ACTIVITY FACTOR FOR THE CLOCK IS >1 BECAUSE IT TOGGLES TWICE PER CYCLE!

$$E_{DATA} = \alpha \tfrac{1}{2} C V_{dd}^2 = (0.1)\tfrac{1}{2}(52\text{C})\frac{0.5\text{ff}}{\text{C}}(1\text{V})^2 = 1.3\,\text{fJ}$$

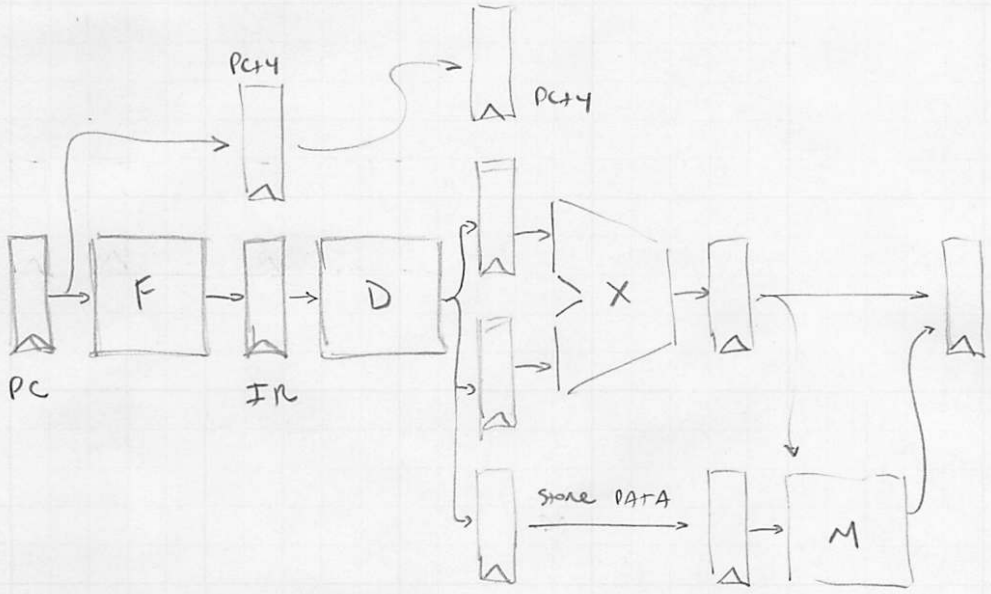
$$P_{DATA} = \alpha f \tfrac{1}{2} C V_{dd}^2 = (0.1)(0.5\times10^9)\tfrac{1}{2}(52\text{C})\frac{0.5\text{ff}}{\text{C}}(1\text{V})^2 = 0.65\,\mu\text{W}$$

$$E_{clock} = \alpha \tfrac{1}{2} C V_{dd}^2 = (2)\tfrac{1}{2}(20\text{C})\frac{0.5\text{ff}}{\text{C}}(1\text{V})^2 = 10\,\text{fJ}$$

$$P_{clock} = \alpha f \tfrac{1}{2} C V_{dd}^2 = (2)(0.5\times10^9)\tfrac{1}{2}(20\text{C})\frac{0.5\text{ff}}{\text{C}}(1\text{V})^2 = 5.\,\mu\text{W}$$

ONCE WE FACTOR IN ACTIVITY FACTOR AND CLOCK FREQUENCY, CLOCK ENERGY/POWER IS SIGNIFICANTLY HIGHER THAN DATA ENERGY

LET'S ESTIMATE THE TOTAL DATA/CLOCK ENERGY FOR A SIMPLE PROCESSOR. HOW MANY FLIPFLOPS IN THE DATAPATH?



ROUGHLY 10 PIPELINE REGISTERS, BUT LET'S ROUND UP TO 16 TO ACCOUNT FOR OTHER STATE IN THE CONTROL UNIT, MEMORY INTERFACE, EXCEPTION PC, ETC. EACH REGISTER IS 32 BITS SO WE MIGHT ESTIMATE A TOTAL OF AROUND

$$16 \times 32 = 512 \text{ bits}$$

NOTE THAT THIS IGNORES THE REGISTER FILE WHICH IS DEFINITELY NOT NEGLIGIBLE

WITH 512 BITS THE TOTAL DATA/CLOCK ENERGY/POWER IS

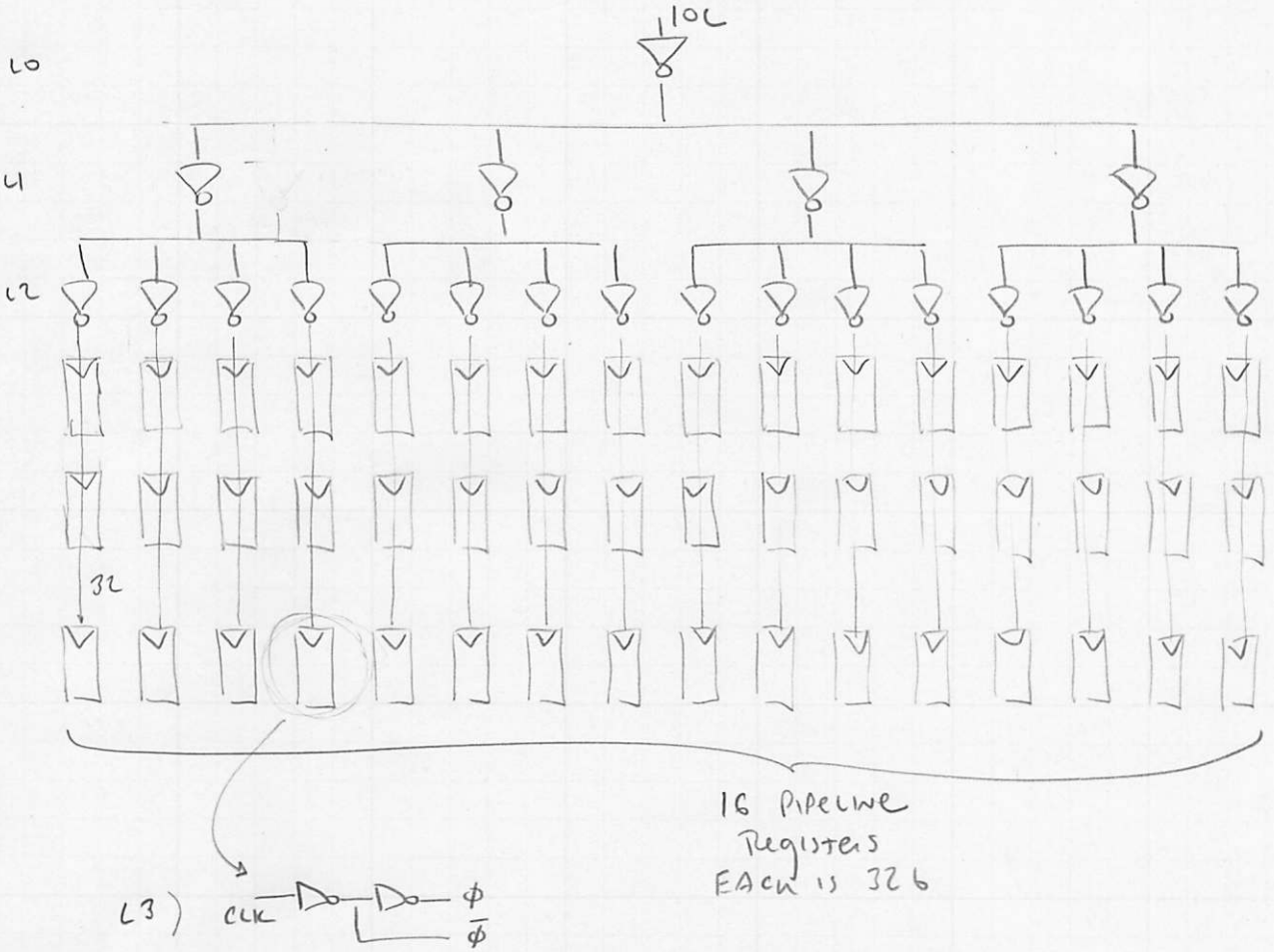
$$E_{DATA,TOT} = 512 \cdot 1.3fJ = 666\,fJ$$

$$P_{DATA,TOT} = 512 \cdot 0.65\mu W = 333\,\mu W$$

$$E_{CLOCK,TOT} = 512 \cdot 10fJ = 5.1\,pJ$$

$$P_{CLOCK,TOT} = 512 \cdot 5\mu W = 2.6\,mW$$

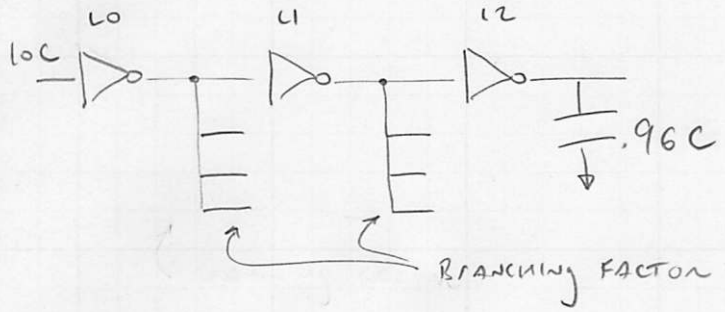
THIS IGNORES THE CLOCK TREE. LET'S TRY AND ESTIMATE THE POWER CONSUMED IN THE CLOCK TREE



EACH ONE BIT FLIP-FLOP ADDS 3C LOAD TO LAST INVERTER IN THE CLOCK TREE, FOR A TOTAL CLOCK LOAD OF $3 \times 32 = 96C$

HOW SHOULD WE SIZE THE INVERTERS IN THE CLOCK TREE TO REDUCE DELAY? NOTE THAT SKEW IS MUCH MORE IMPORTANT THAN ABSOLUTE DELAY BUT WE STILL NEED A RELATIVELY FAST TREE TO AVOID VERY BAD SLEW RATES.

we can use logical effort to size the clock tree



$F = GBH = 1(4\times4)\,96/10 = 153.6$ $\quad \log_4 F = 3.6$

$\hat{f} = \sqrt[3]{F} = 5.4$

$\hat{D} = 3(5.4) + 3 = 19.2\,\tau$

so optimal number of stages for inverter chain with path effort ~154 is 3-4 stages looks like our 3 stage clock tree is in the right ball park

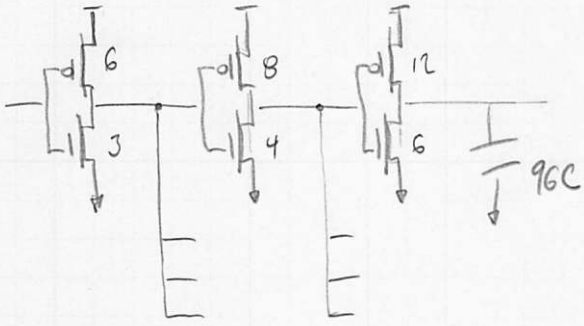
$$C_{in} = \left(\frac{g_i}{\hat{f}}\right) C_{out}$$

$C_{in,L2} = \frac{1}{5.4}\,96C = 17.8C$

$C_{in,L1} = \frac{1}{5.4}\,17.8C \times 4 = 13.2C$

$C_{in,L0} = \frac{1}{5.4}\,13.2C \times 4 = 9.8C$

let's round to nearest even multiple of minimum inverter



TOTAL SWITCHED CAP?

$C_{sw,L2} = 18C \times 2 \times 16 = 576C$

$C_{sw,L1} = 12C \times 2 \times 4 = 96C$

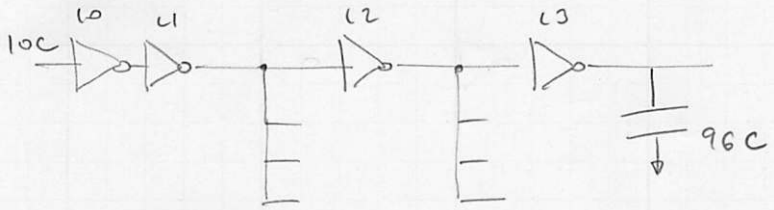
$C_{sw,L0} = 9C \times 2 \times 1 = 18C$

→ THIS FACTOR OF TWO IS BECAUSE WHEN WE CALCULATE TOTAL SWITCHED CAPACITANCE WE CARE ABOUT THE GATE CAP AND THE PARASITIC CAP.

TOTAL C_{sw} for 3 stage tree is 690C

$P_{clk,tree} = \alpha f \frac{1}{2} C V_{DD}^2 = 2(0.5\times10^9)\frac{1}{2}\,690C\left(\frac{0.5fF}{C}\right)(1V)^2 = 173\,\mu W$

WHAT ABOUT A FOUR-LEVEL CLOCK TREE?



$F = GBH = 1\,(4\times4)\,96/10 = 153.6$

$\hat{f} = \sqrt[4]{F} = 3.52$

$\hat{D} = 4\,(3.52) + 4 = \underline{18.1\gamma}$ COMPARED TO 19.2 WITH ONLY 3 LEVELS

SO OUR FOUR-LEVEL CLOCK TREE IS ~5% FASTER BUT AT WHAT COST?

$C_{IN,L3}$	$= (1/3.52)\ 96C$	$= 27.3C$	$27C$
$C_{IN,L2}$	$= (1/3.52)\ 27.3C\times4$	$= 31C$	$30C$
$C_{IN,L1}$	$= (1/3.52)\ 31C\times4$	$= 35C$	$36C$
$C_{IN,L\phi}$	$= (1/3.52)\ 35C$	$= 9.94C$	$9C$

ROUND TO NEAREST EVEN MULTIPLE OF MINIMUM SIZED INVERTER

TOTAL SWITCHED CAP?

$C_{SW,L3} = 27C \times 2 \times 16 = 864C$

$C_{SW,L2} = 30C \times 2 \times 4 = 240C$

$C_{SW,L1} = 36C \times 2 \times 1 = 72C$

$C_{SW,L0} = 9C \times 2 \times 1 = 18C$

TOTAL C_{SW} for 4 STAGE TREE IS 1194 C

SO 4 STAGE TREE IS 5% FASTER BUT ~1.7X MORE ENERGY/POWER/AREA
LET'S STICK WITH THE THREE STAGE!

SO IN SUMMARY

$P_{DATA} = 333\,\mu W$

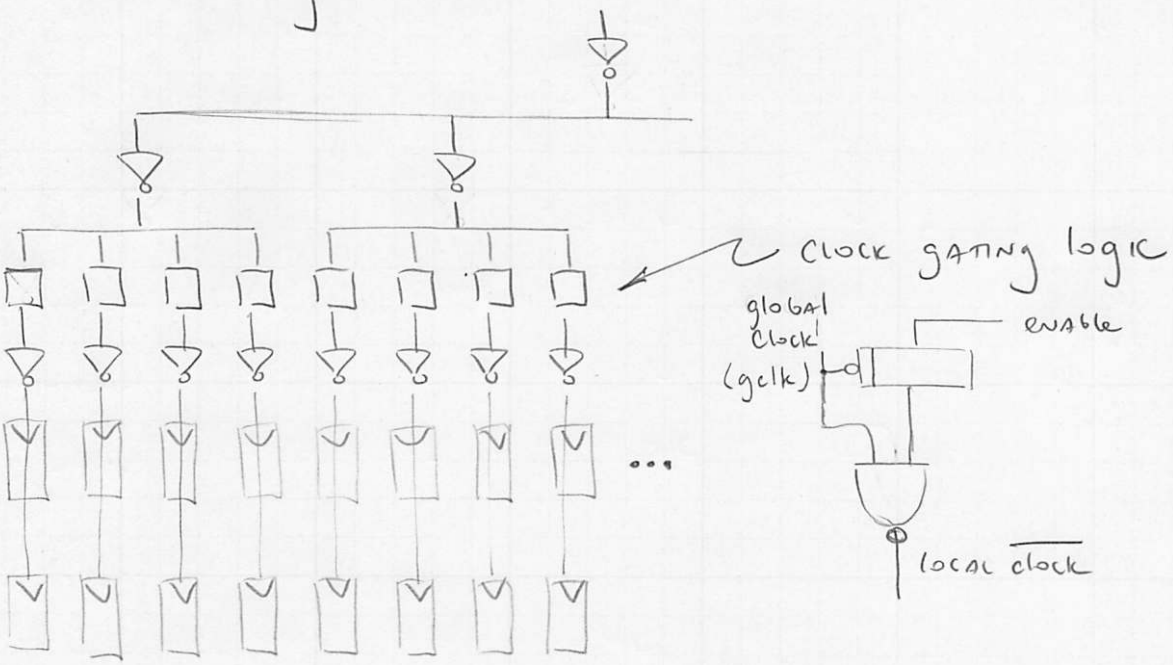
$P_{CLOCK} = 7.6\,mW$

$P_{CLOCK,TREE} = 173\,\mu W$

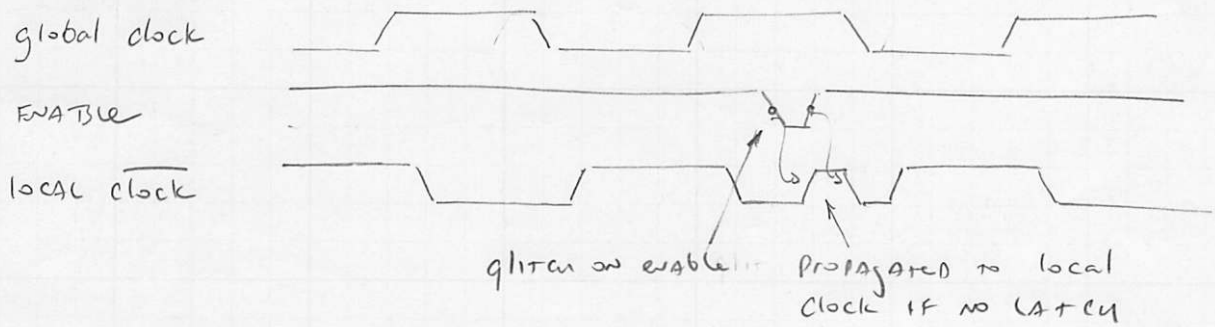
$P_{TOT} = 3.1\,mW$

SO ALMOST 84% of TOTAL DYNAMIC POWER IS SPENT IN THE LOCAL "INTRA-cell" clock logic!

CAN CLOCK GATING HELP?



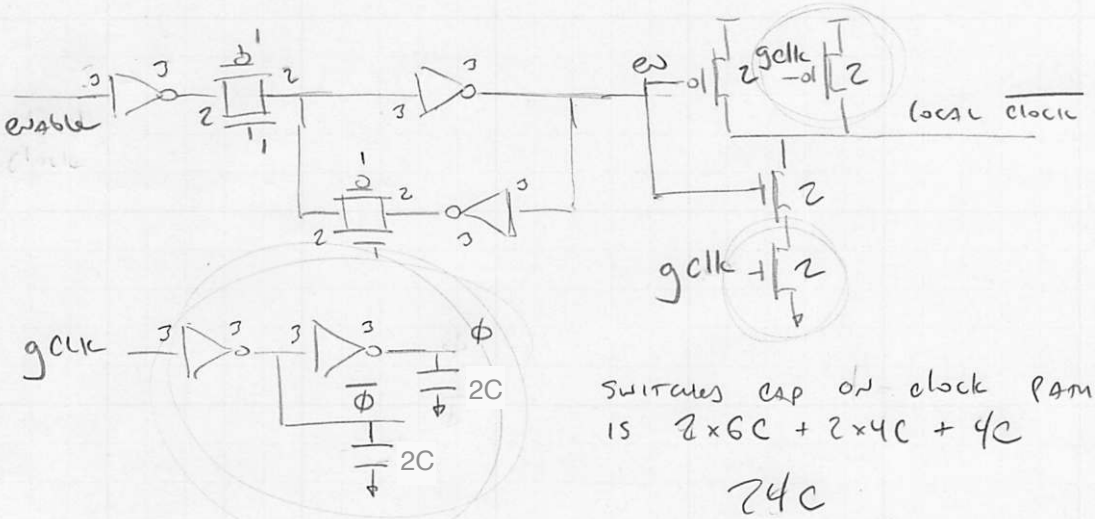
WHY DO WE NEED THE LATCH? AVOID glitching ON enable SIGNAL FROM propagating TO local clock



THE FACT THAT local clock is complement of global clock is NOT A problem WITH OUR TRANSMISSION GATE FLIPFLOPS

Clock gating reduces activity on L2 of clock tree and local intra-cell clocking logic but also adds extra switches cap when Register is enables.

Let's estimate extra switches cap due to gating cell on clock path when register is enables continuously



switches cap on clock path is $2 \times 6C + 2 \times 4C + 4C$

$$24C$$

What about when disables? Assume enable signal is continuously set to zero. It is still 24C, but of course the local clock does not toggle.

The local clock is 576 C for final inverter in clock tree plus an additional $512 \times 20C = 10,240$ C for all of the local intra-cell clock logic.

	Data	intra cell clock	L0, L1 clk tree	L2 clk tree	gating logic	total
no clk gating	333 μW	26 mW	28.5 μW	144 μW	N/A	3.1 mW
w/ clk gating + enables	333 μW	2.6 mW	28.5 μW	144 μW	96 μW	3.2 mW
w/ clk gating + Disables	333 μW	0 mW	28.5 μW	0 μW	96 μW	458 μW

So in the worst case where we add clock gating but we never actually gate any registers, we add a power overhead of ~3%

In the best case when every register is gated we reduce the total power for the pipeline registers by 85%