ECE4740: Digital VLSI Design

Lecture 19: Dynamic latches/flip-flops

Recap
Timing, flip-flops, and latches

## Common flip-flop and latch symbols


rising-edge triggered FF

falling-edge triggered FF

positive latch 1-transparent 0-hold

negative latch 0-transparent 1-hold

- Real-world flip-flops (and latches) may have more inputs and outputs, such as
- Reset in, enable in, scan in, and !Q out


## Positive latch: transparent if CLK=1



## Positive-edge triggered MS flip-flop



CLK $=0 \rightarrow$ master transparent; slave hold

## Positive-edge triggered MS flip-flop



CLK=1 $\rightarrow$ master hold; slave transparent

## Setup and hold times



## Non-ideal clocks: clock skew



## 1-1 overlap is dangerous



- Direct path from D to Q during short time when both CLK and !CLK are high
- Happens during 1-1 overlap


## 1-1 overlap is dangerous (cont'd)



- Both B and D are driving A when CLK and !CLK are both high (1-1 overlap)


## Generating a non-1-1-overlapping clock

- To avoid overlapping clocks 1-1 we need
- tools for accurate timing analysis OR
- non-1-1-overlapping clock signals
- One can use SR-latch to generate such clocks


Building sequential logic with fewer transistors

## Dynamic latches and flip-flops

## Static vs. dynamic storage cells

- Static cells use bistable element with feedback (regeneration)
- Preserve state as long as power is on
- Static storage is preferred when updates are infrequent (clock gating etc.)
- Dynamic storage on parasitic capacitors
- Preserve state only for milliseconds
- Dynamic storage cells are usually smaller, achieve higher speed and consume lower power


## Dynamic edge-triggered flip-flop


master transparent slave hold


## Dynamic ET flip-flop (cont'd)



- Requires only 8 transistors; clock load = 4
- Dynamic nodes need periodical refresh


## Issue 1: race conditions


data must be stable
during high phase

## Solution: non-overlapping clocks



## Issue 2: robustness

- Dynamic flip-flops suffer from
- Coupling between signal nets and internal storage nodes (can destroy FF state)
- Leakage currents cause state to leak with time
- Solution: pseudostatic FF



## A clock-skew insensitive approach

The C²MOS register

C²MOS (clocked CMOS) ET FF

master transparent
slave hold


## C²MOS FF: 0-0 overlap



## C²MOS FF: 1-1 overlap



1-1 overlap constraint
(Slope matters: transient response)


For high-throughput designs
Pipelining \& retiming

## Consider the timing of this circuit



- Critical path:

$$
\mathrm{T}_{\text {min }}=\mathrm{t}_{\mathrm{pd}, \mathrm{ff}}+_{\mathrm{tpd}, \mathrm{add}}+{ }_{\mathrm{tpd}, \mathrm{abs}}+\mathrm{t}_{\mathrm{pd}, \log }+\mathrm{t}_{\text {su,ff }}
$$

## Pipelining reduces critical path



- Insert pipeline registers (flip-flops)
- Shortens critical path!

$$
\mathrm{T}_{\text {pipe, min }}=\mathrm{t}_{\mathrm{pd}, \mathrm{ff}}+\max \left\{\mathrm{t}_{\mathrm{pd}, \text { add }}, \mathrm{t}_{\mathrm{pd}, \mathrm{abs},}, \mathrm{t}_{\mathrm{pd}, \mathrm{log}}\right\}+\mathrm{t}_{\mathrm{su}, \mathrm{ff}}
$$

## Pipelining



## Pipelining (cont'd)



## Pipelining (cont'd)



## Pipelining improves throughput!



- Processes 1 data item per clock cycle at higher $f_{\max }$ $\rightarrow$ higher throughput (time per data item $\mathrm{T}_{\text {min,pipe }}$ )
- Ideally: $T_{\text {min, ,pipe }}=t_{\text {pd, ff }}+t_{\text {pd, logic }} / N+t_{\text {su,ff }}$ with $N$ stages
- Throughput limit: $T_{\text {min, pipe }} \geq t_{\text {pd,ff }}+\mathrm{t}_{\text {su,ff }}$


## Pipelining introduces latency



- Latency = \# of cycles for data to propagate from input to output
- Latency $=4$ (four rising clock edges)

- If feedback path is present, latency will reduce throughput (circuit has to wait for data)
- Problem in processors and application specific integrated circuits (data dependencies)


## Solution: Pipeline interleaving


reduces
throughput by $2 x$

- Idea: Process independent problems in an interleaved manner in the same hardware
even cycles

odd cycles



## Pipelining using $C^{2} \mathrm{MOS}$



- Circuit is race-condition free (NORA) if functions $F$ and $G$ are non-inverting!


## Your turn: pipeline a MAC unit

 multiply-accumulate (MAC) unit: $D=B^{*} C+A$

- What is the max. clock frequency?
- Where is the critical path?
- Insert a single pipeline stage
- What is the max. clock frequency after pipelining?


## Critical path and max. clock freq.



- $T_{\text {min }}=t_{\text {pd,ff }}+t_{\text {pd, mult }}+t_{\text {pd,add }}+t_{\text {su,ff }}=10 \mathrm{~ns}$
- $f_{\max }=100 \mathrm{MHz}$


## Pipelining: max. clock freq. now?



- $T_{\text {min, pipe }}=t_{\text {pd,ff }}+t_{\text {pd,mult }}+t_{\text {su,ff }}=8 n s$
- $f_{\max }=125 \mathrm{MHz}$

