Module

9

ECE M216A

Clocking Methodologies

Prof. Dejan Marković

ee216a@gmail.com

Agenda

- **Clock distribution**
- **Clock gating**
  - Global
  - Local
- **Low-energy flip-flops**
  - Low-swing clock
  - Dual-edge triggered

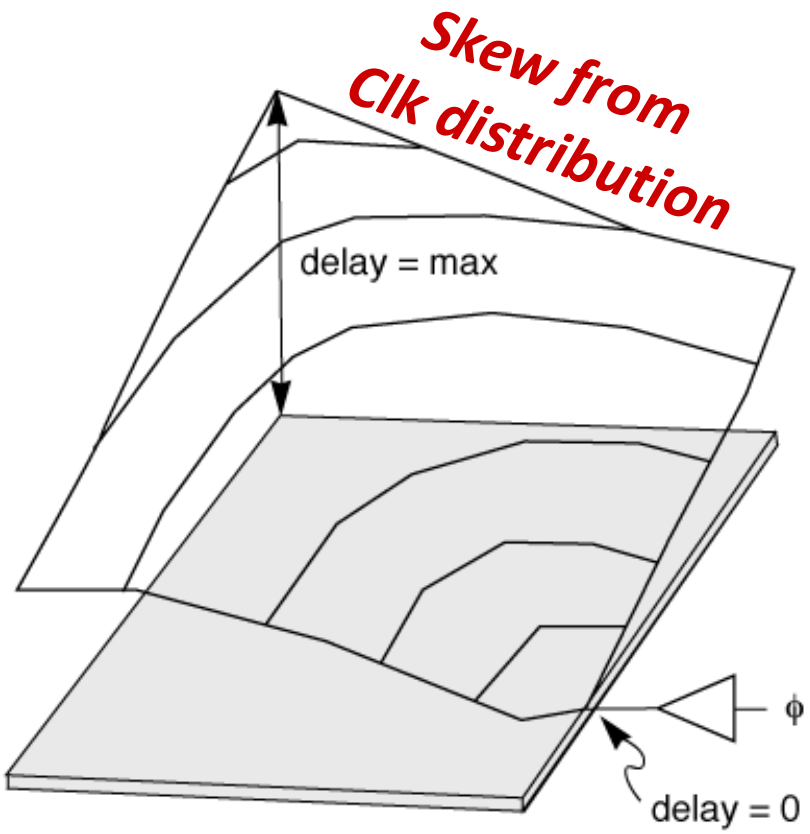
Clock Distribution Goals

- **Deliver clock to all memory elements with acceptable skew**
- **Deliver clock edges with acceptable sharpness**

Clock Distribution

- **Challenging in the design of a large chip**
- **Clocks are generally distributed via wiring trees**
- **Use low-R interconnect to minimize delay**
- **Use multiple drivers to distribute driver requirements**
  - Use optimal sizing principles to design buffers
  - Clock lines can create significant crosstalk

Issues in Clock Distribution Network

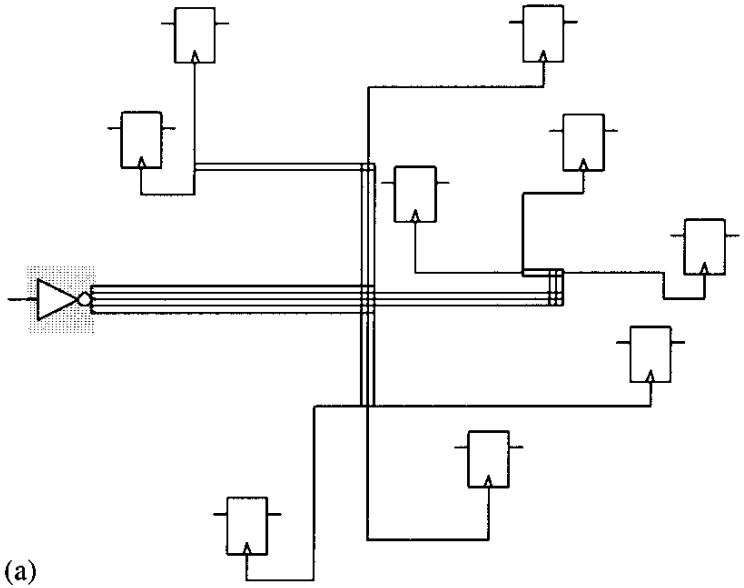


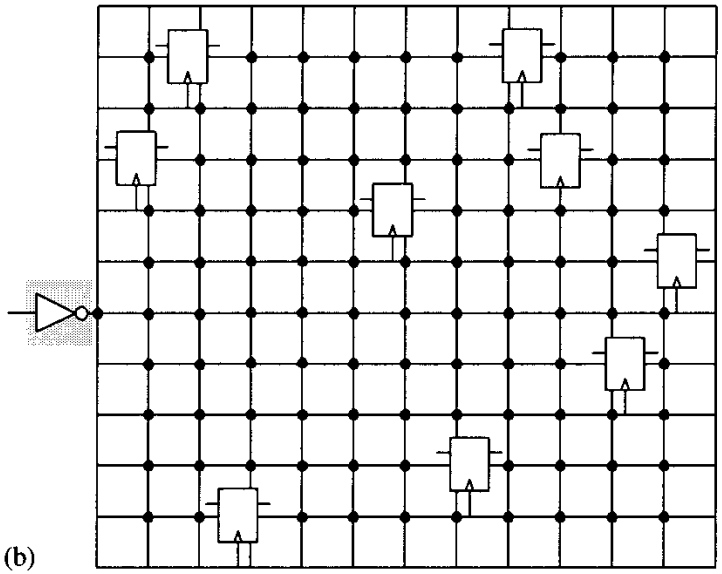
See videos from P. Restle (IBM) on classwiki

- **Skew**
  - Process, voltage, temp.
  - Data dependence
  - Noise coupling
  - Load balancing
- **Power, CV^2f – (no ½ or α)**
  - Clock gating
- **Flexibility/Tunability**
  - Fit into existing design
- **Reliability**
  - Electromigration

Clock Distribution Methods

- **RC-Trees**
  - Less capacitance
  - More accuracy
  - Flexible wiring
- **Grids**
  - Reliable
  - Less data dependency
  - Tunable (late in design)



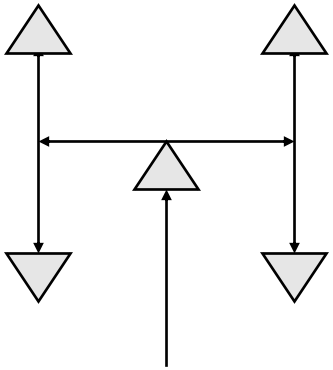


Shown here for final-stage drivers driving F-F loads

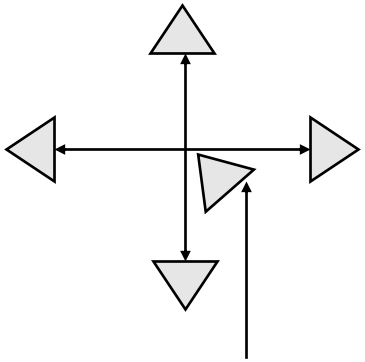
Clock Distribution: RC Trees

Note: only relative skew is important

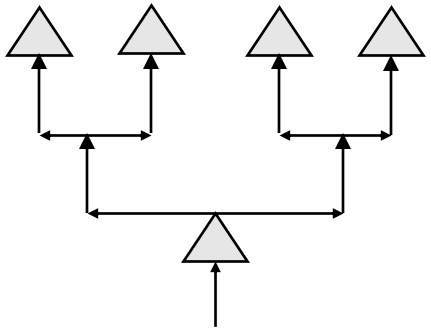
H-Tree



X-Tree

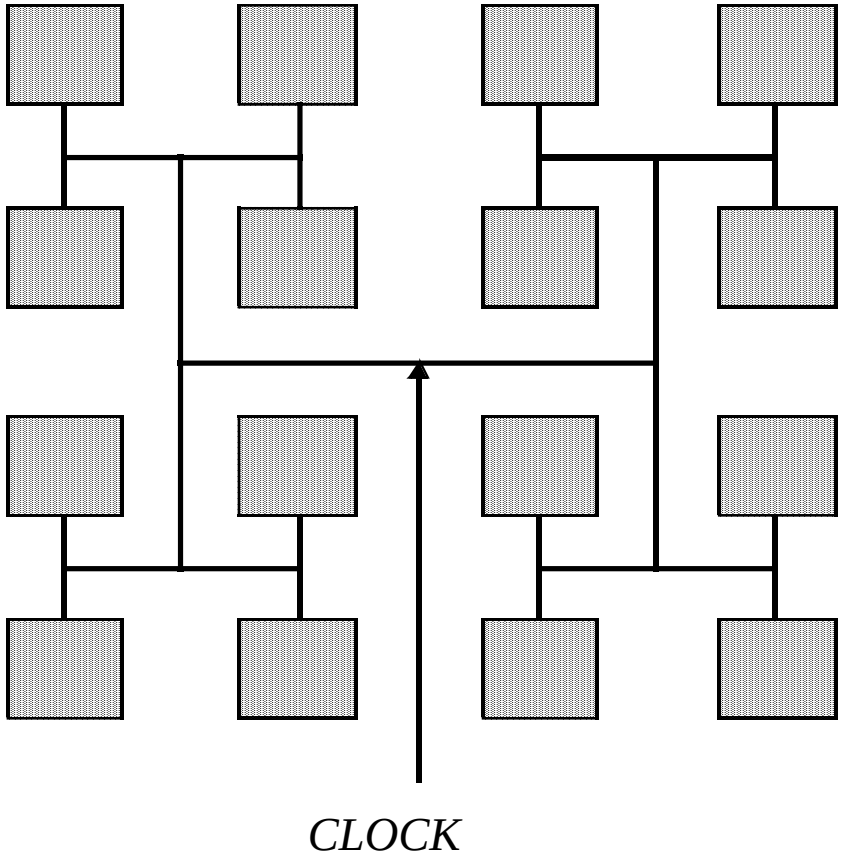


Binary-Tree



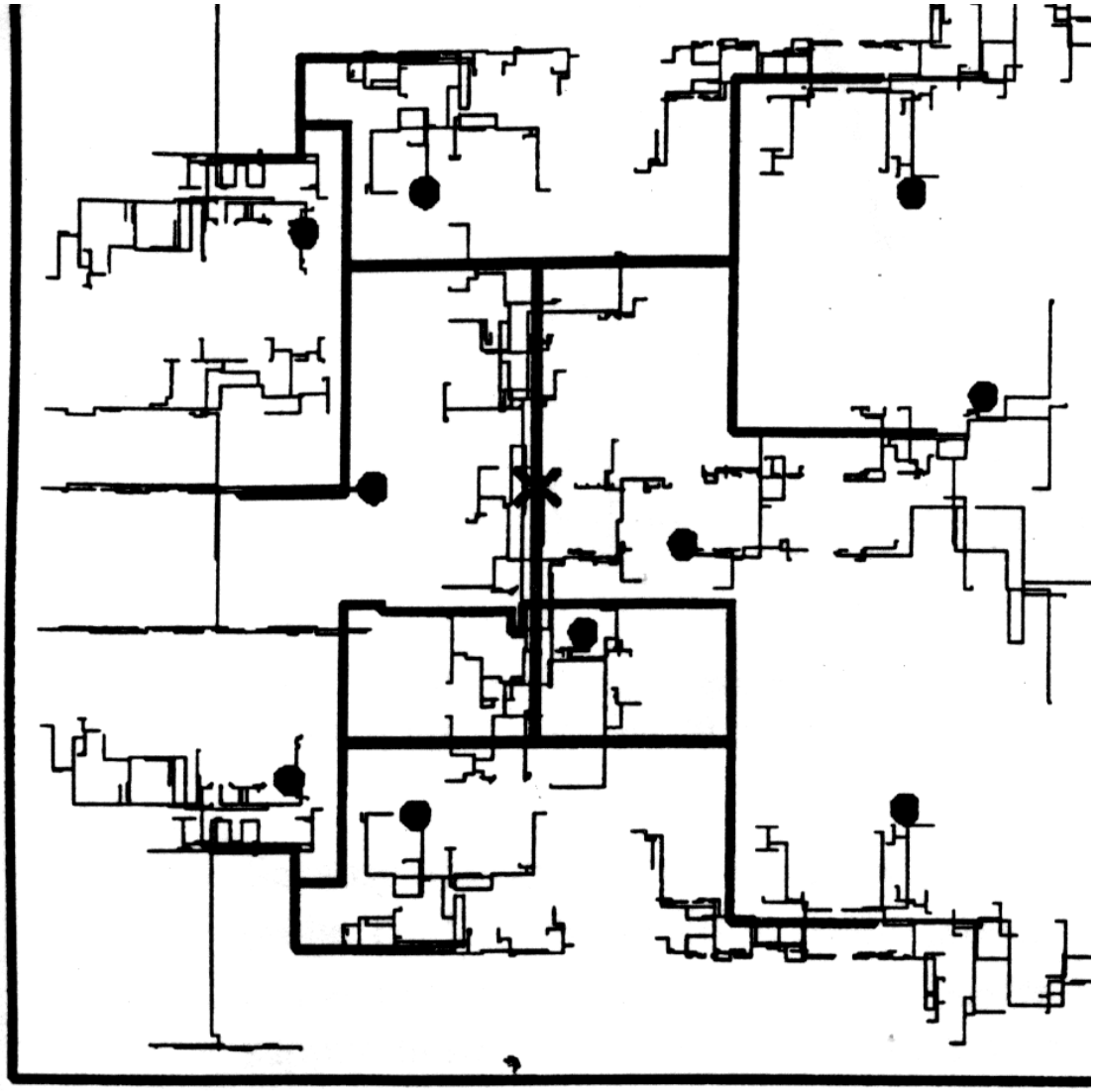
Asymmetric trees that match RCs can be used

H-Tree Network



- Radix-4
- FO = 4
- Sharp edges
- Low delay

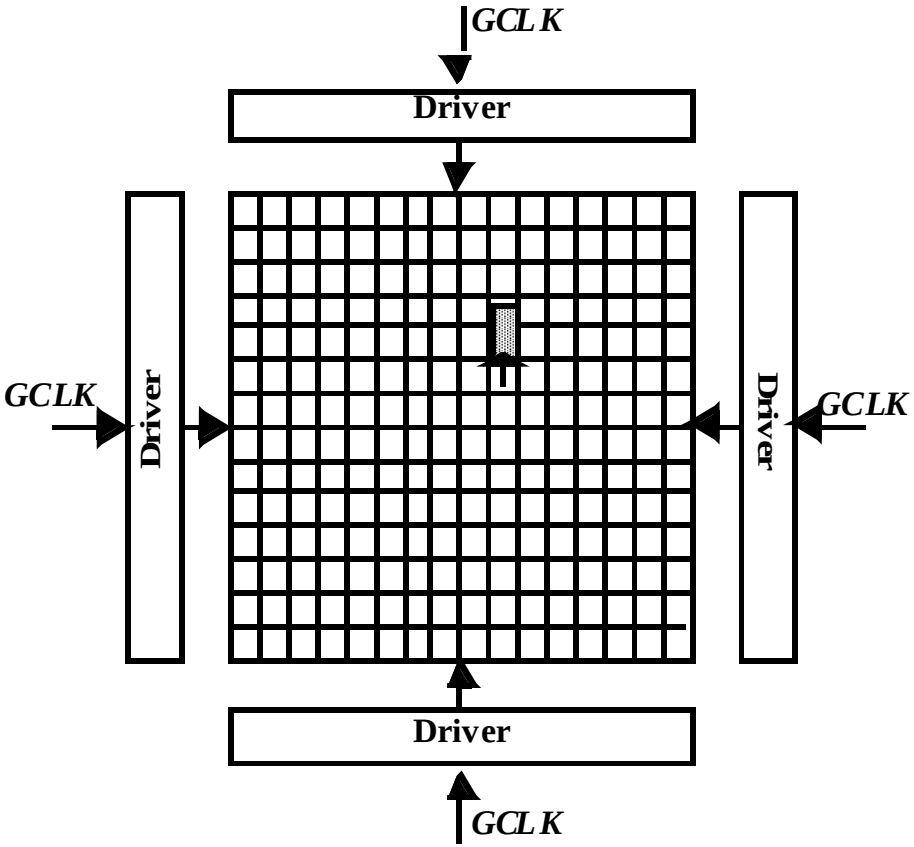
More Realistic H-Tree



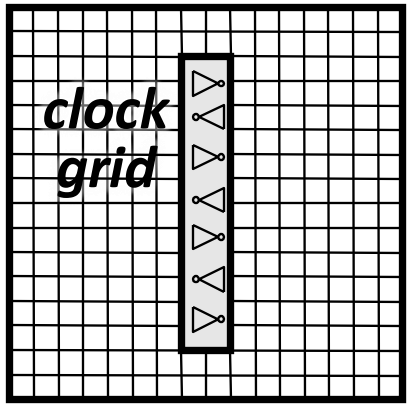
[Restle98]

The Grid System

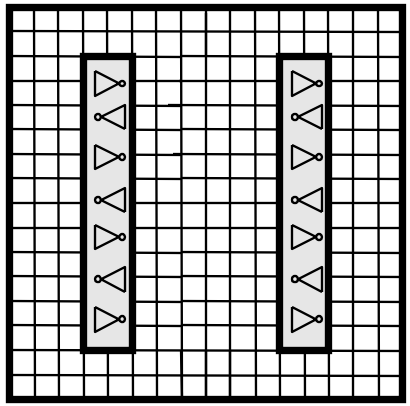
- No RC-matching
- Large power



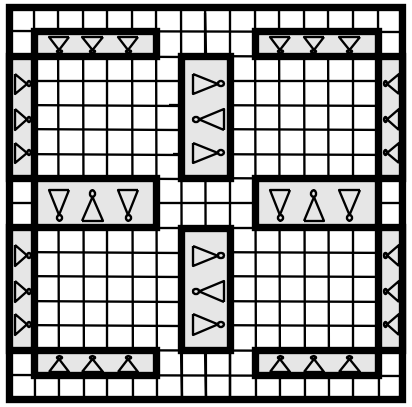
DEC Alpha Examples



21064



21164



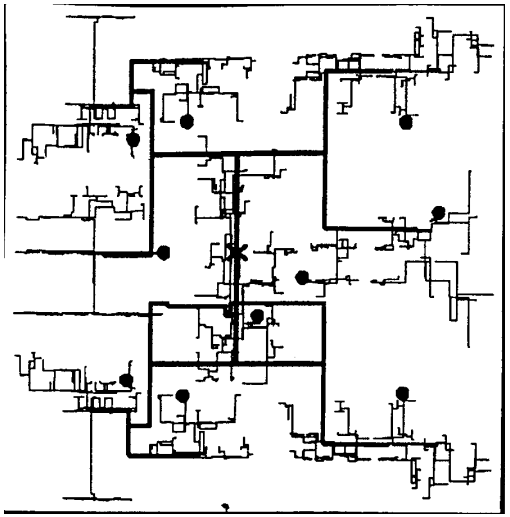
21264

Examples of Distribution

H-Tree

IBM

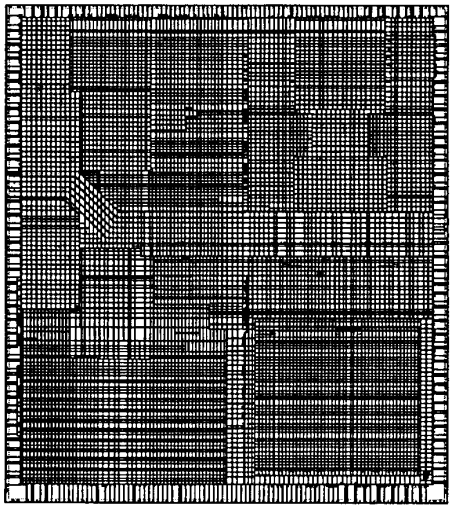
(Asymmetric RC-Tree)



Grids

DEC

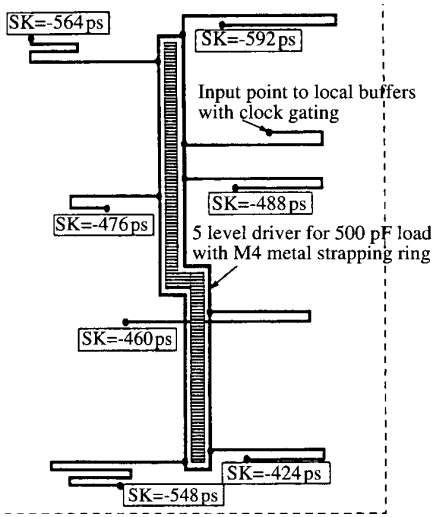
[Alphas]



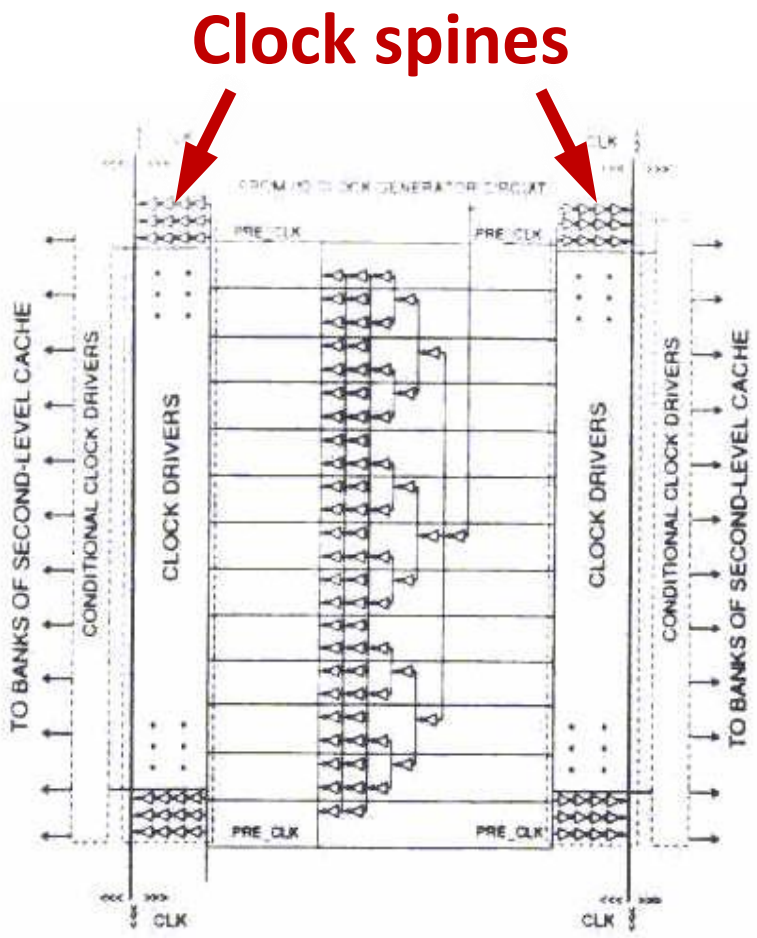
Serpentines

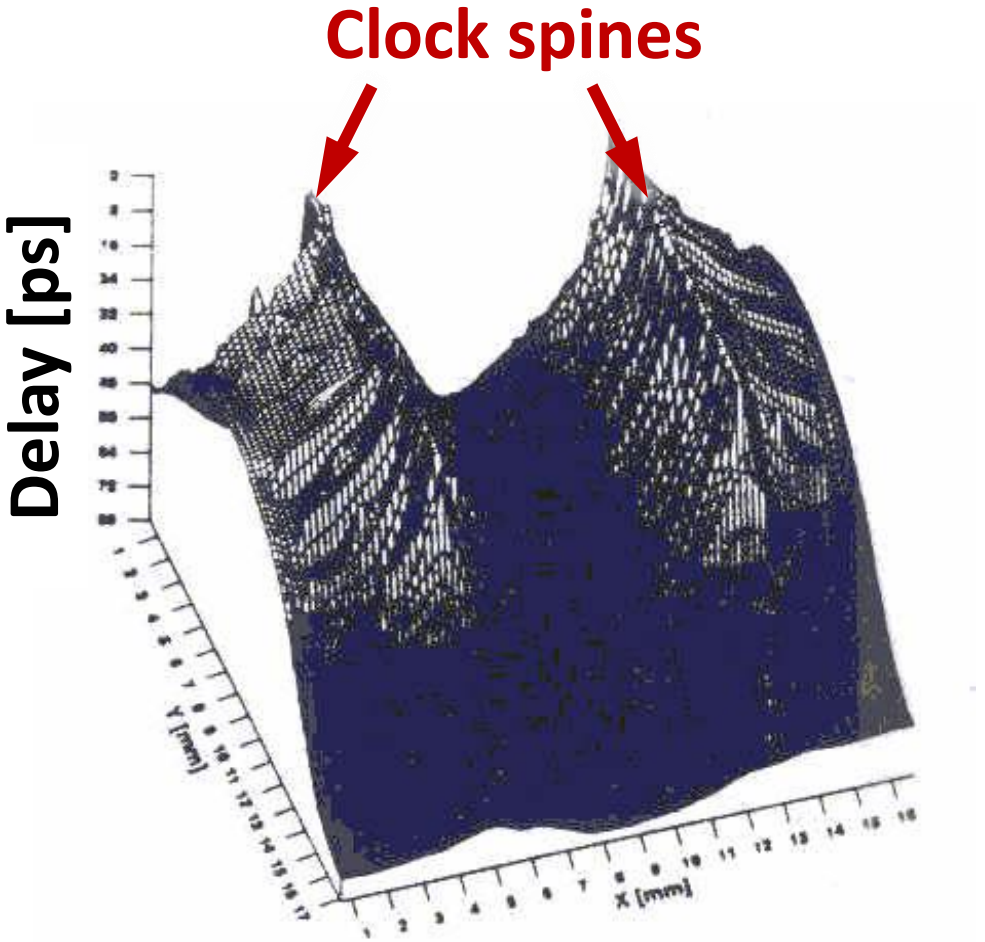
Intel x86

[Young ISSCC97]



Example 9.1: DEC-Alpha 21164

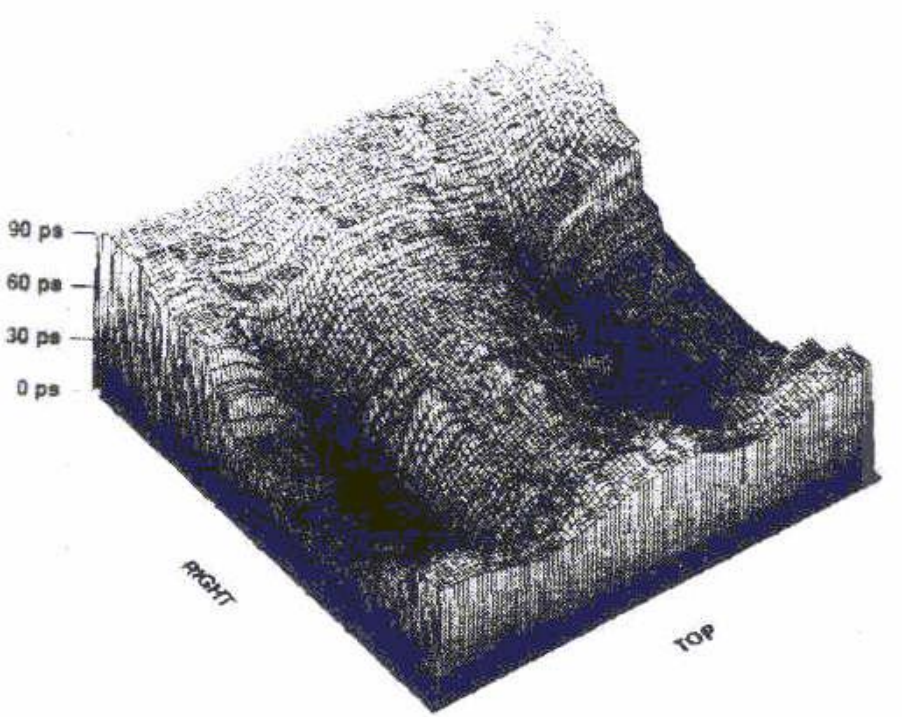




Note: reverse Z-axis

Example 9.2a: DEC-Alpha 21164

Distribution

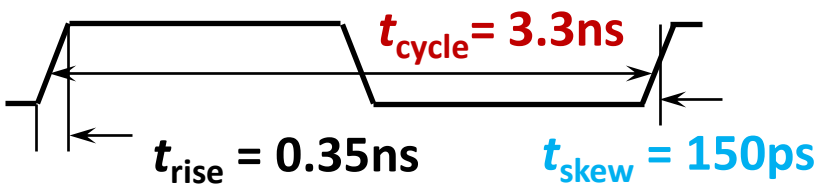


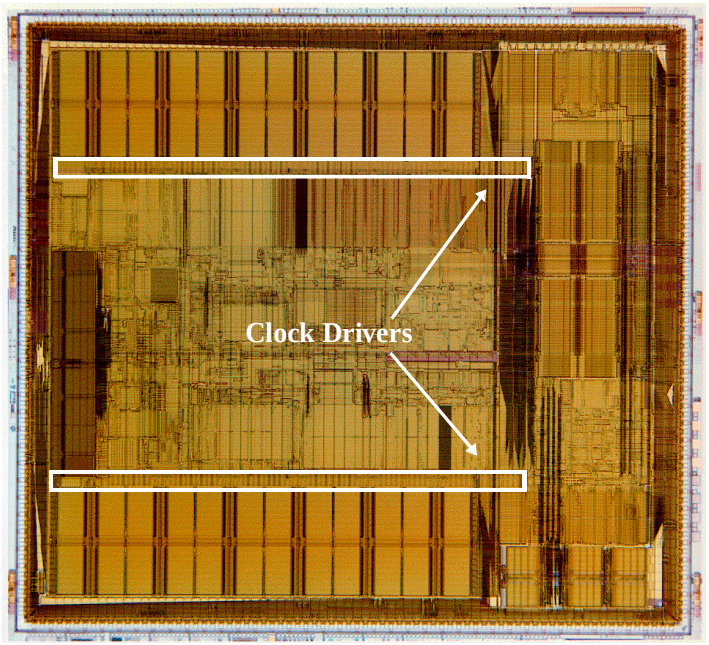
Global distribution

Spine + Grid

Local delays

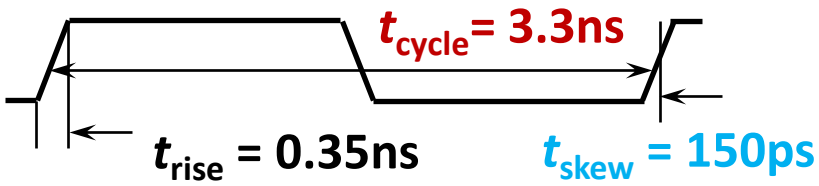
Example 9.2b: EV5 (21164) Clocking (1995)



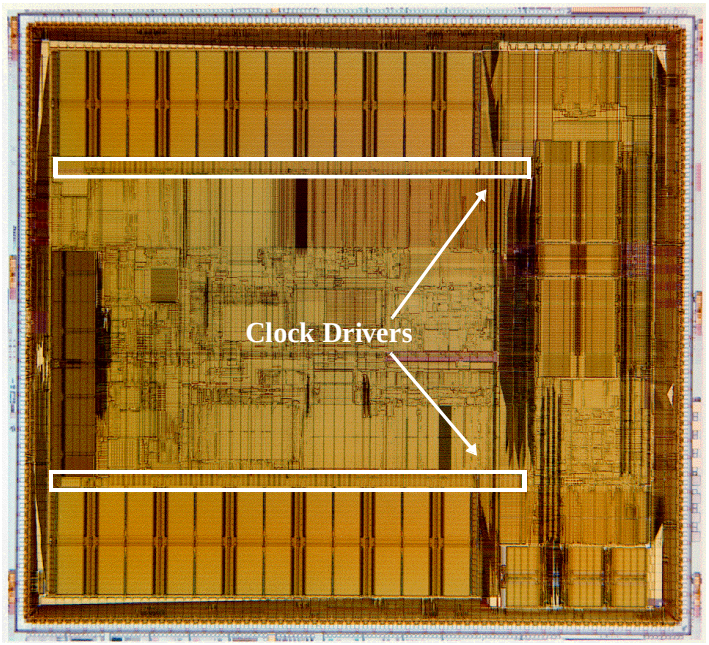


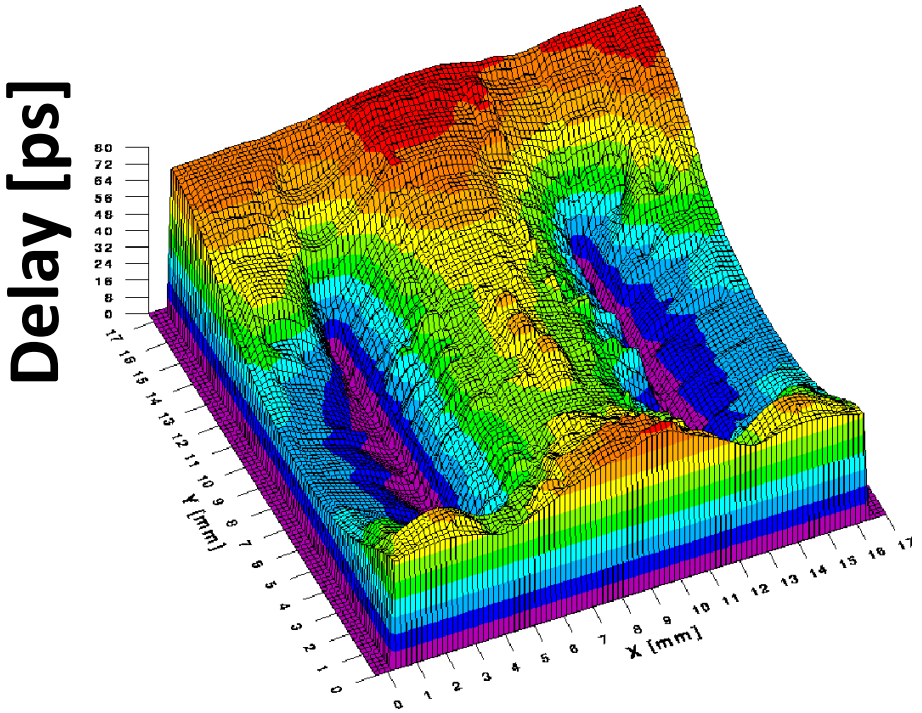
- **Single-phase clocking**
- **2 distributed driver channels**
  - Reduced RC delay/skew
  - Improved thermal distribution
  - **3.75 nF** clock load
  - **58 cm** final driver width
- **Local inverters for latching**
- **Conditional clocks in caches to reduce power**
- **More complex race checking**
- **Device variation**

Example 9.2c: EV5 (21164) Clocking (1995)

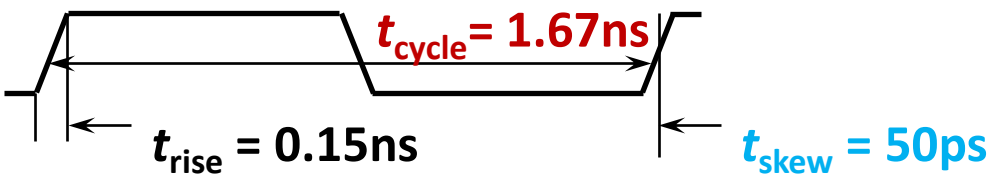


300 MHz
0.5-µm CMOS

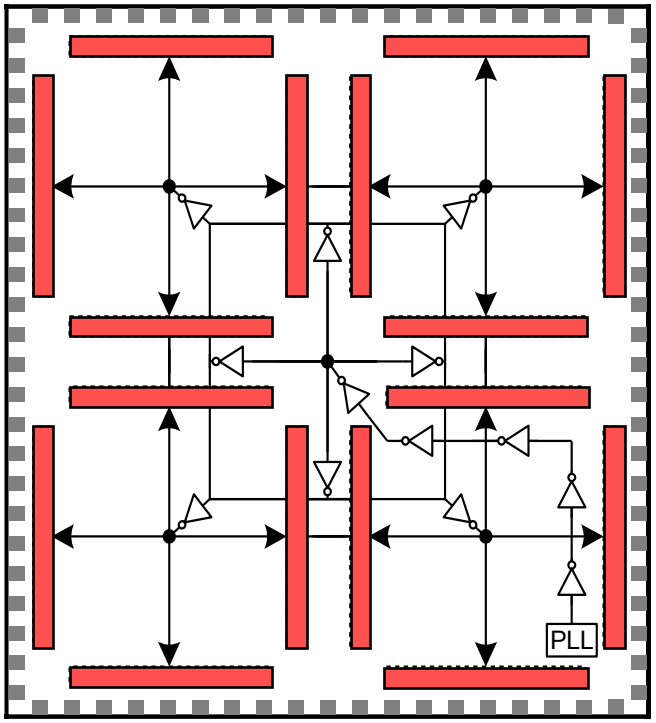




Example 9.3a: EV6 (21264) Clocking (1998)

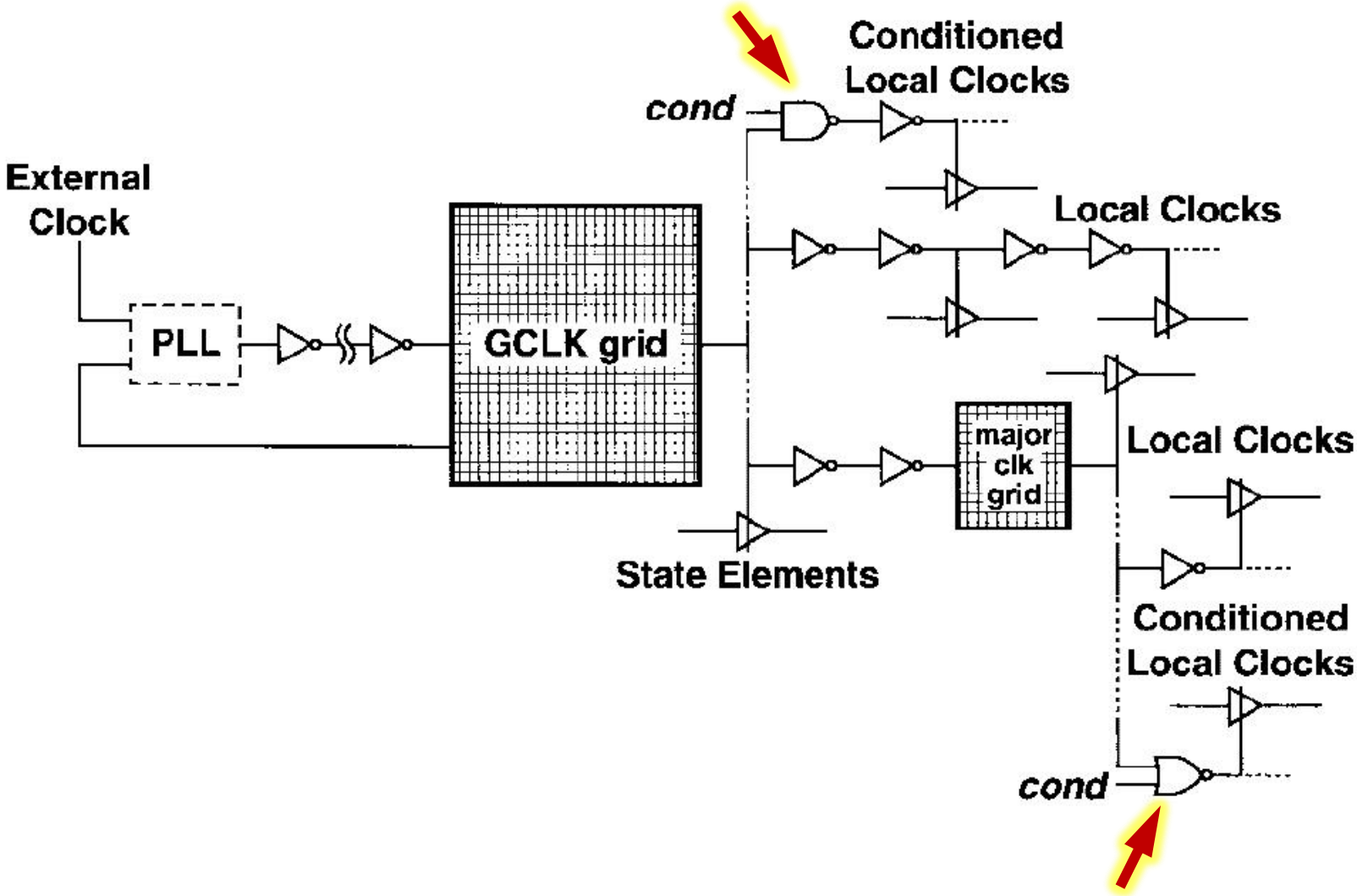


600 MHz
0.35-µm CMOS

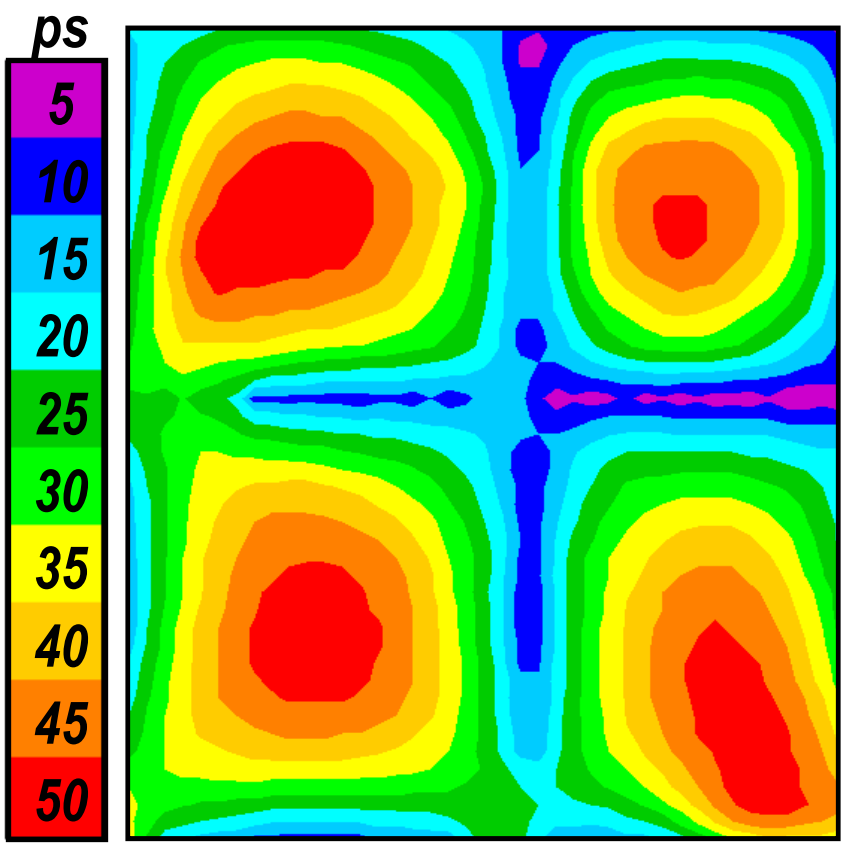


- **Multiple conditional buffered clocks**
  - **2.8 nF** clock load
  - **40 cm** final driver width
- **Reduced load/skew**
- **Multiple clocks complicate race checking**

Example 9.3b: EV6 (21264) Clocking (1998)

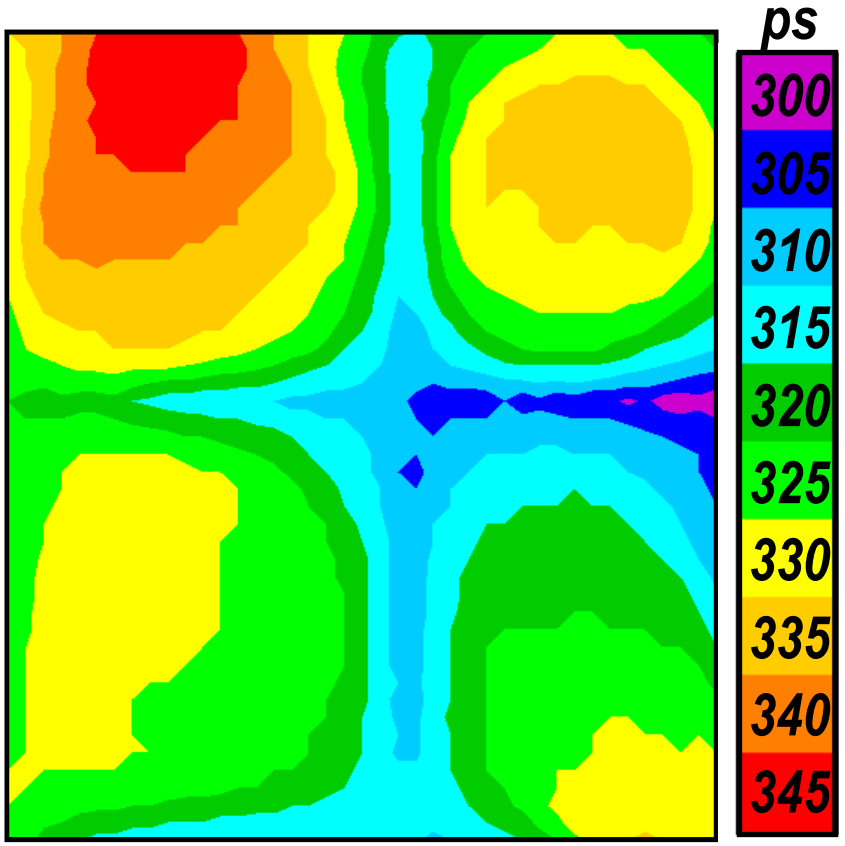


Example 9.3c: EV6 Clocking Results



GCLK Skew

(at $V_{DD}/2$ crossings)



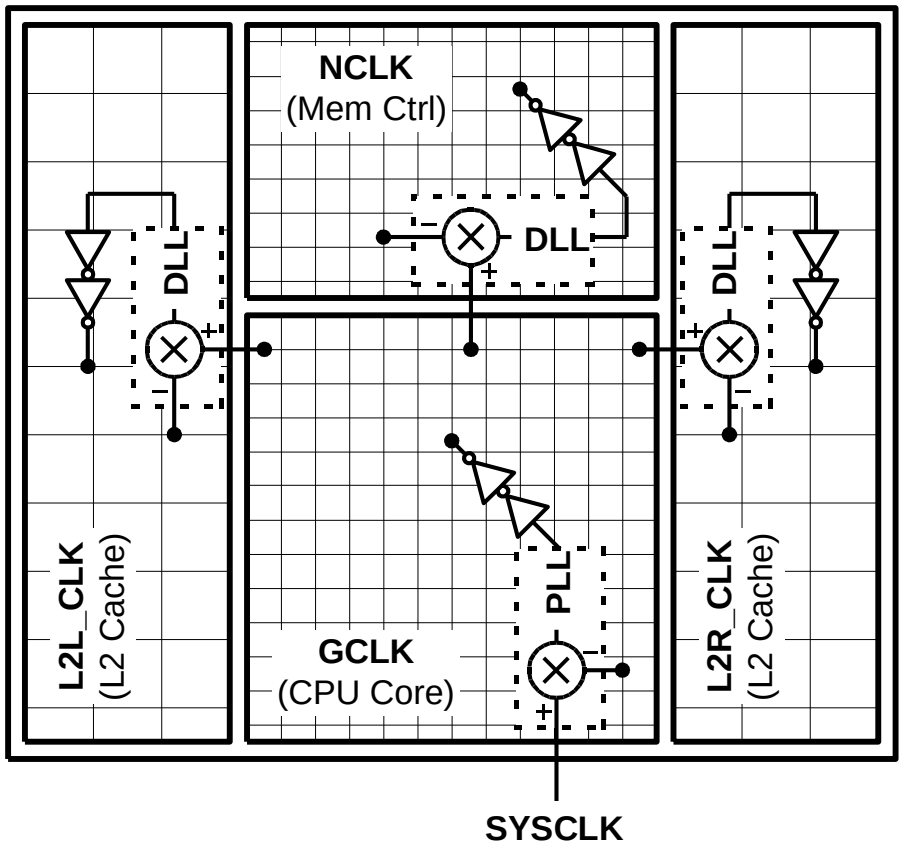
GCLK Rise Times

(20% to 80% extrapolated to 0% to 100%)

Example 9.4: EV7 Clocking (2002)

152 million transistors, 15/137 logic/memory

Active Skew Management and Multiple Clock Domains



- Widely dispersed drivers
- DLLs compensate static and low-freq variation
- Divides design and verification effort
- Extra work in DLL design, verification
- Tailored clocks

Alpha Processors Case Study

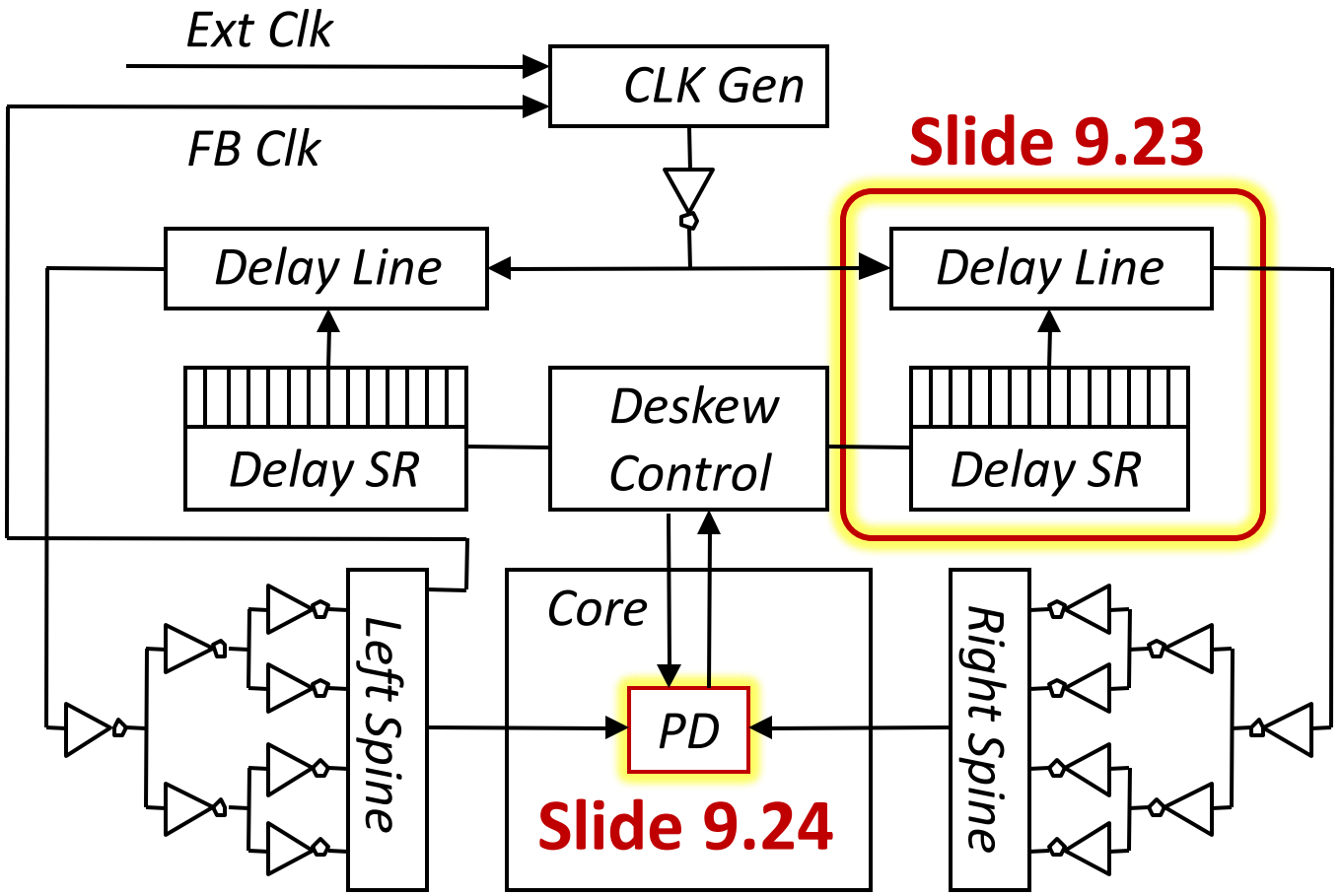
Processor	EV4	EV5	EV6	EV7
Technology	0.75µm	0.5µm	0.35µm	0.18m
Year	1992	1995	1998	2002
Clock rate	200 MHz	300 MHz	600 MHz	1.2 GHz
Load	3.25 nF (40% Power)	3.75 nF (40% Power)	2.8 nF	N/A
Driver	35 cm	58 cm	40 cm	N/A
Cocking	Single global Clk driver, 5 levels of buffering	One central, two side Clk drivers	Clk grid, 4 window panes, hierarchical, gated Clk domains	Multiple Clk domains, DLLs

Intel Processors

Processor	Pentium® II	Pentium® III	Pentium® 4
MPR Issue	June 1997	April 2000	Dec 2001
Clock Speed	**266 MHz**	**1GHz**	**2GHz**
Pipeline Stages	12/14	12/14	22/24
Transistors	**7.5M**	**24M**	**42M**
Cache (I/D/L2)	16k/16K/-	16K/16K/256K	12K/8K/256K
Die Size	203mm^2	106mm^2	217mm^2
IC Process	0.25µm, 4M	0.18µm, 6M	0.18µm, 6M
Max Power	**27W**	**23W**	**67W**

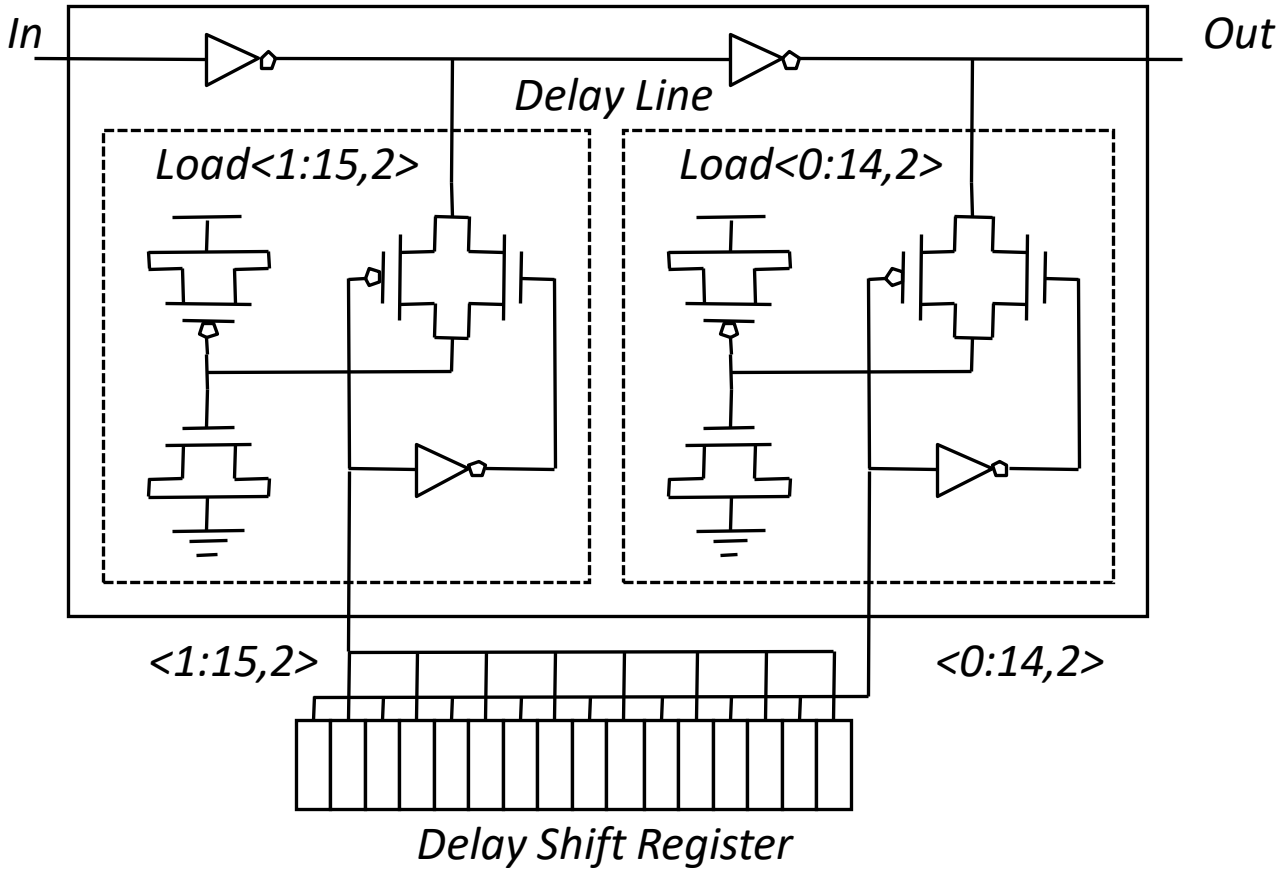
- **Balancing skew using RC trees becoming less effective**
- **Insertion delay 7-8 FO4 due to increased die**
- **Clock skew control harder: increasing PVT variations**
- **Active deskewing circuits**

IA-32 Pentium Pro: Active Deskewing



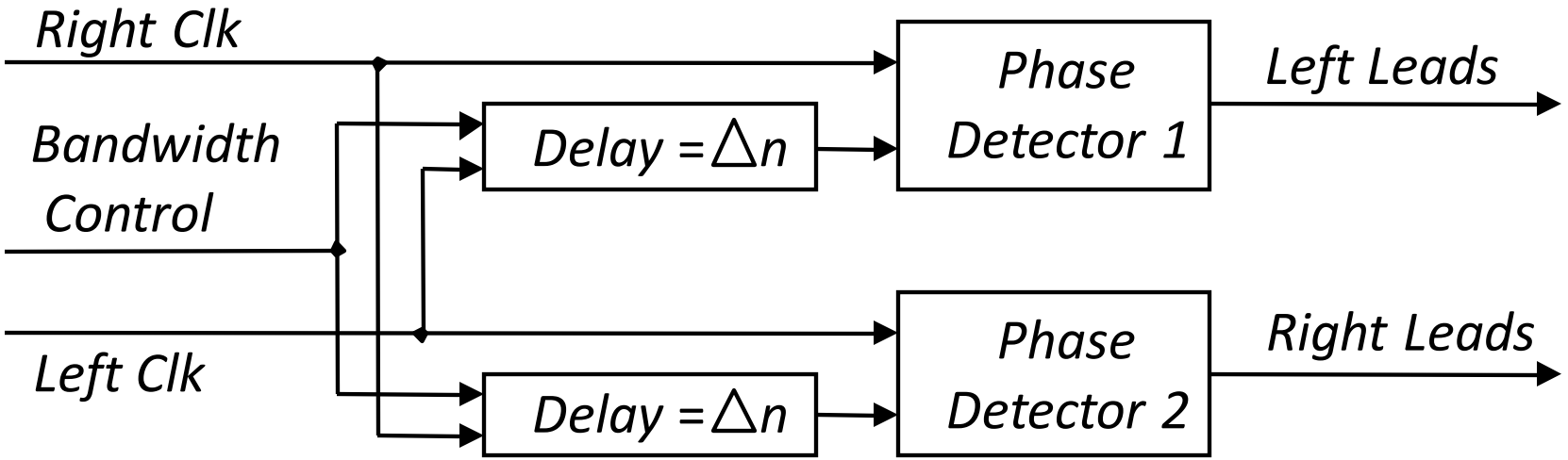
[Geannopoulos and Dai 1998]

IA-32 Pentium Pro: Delay Shift Register



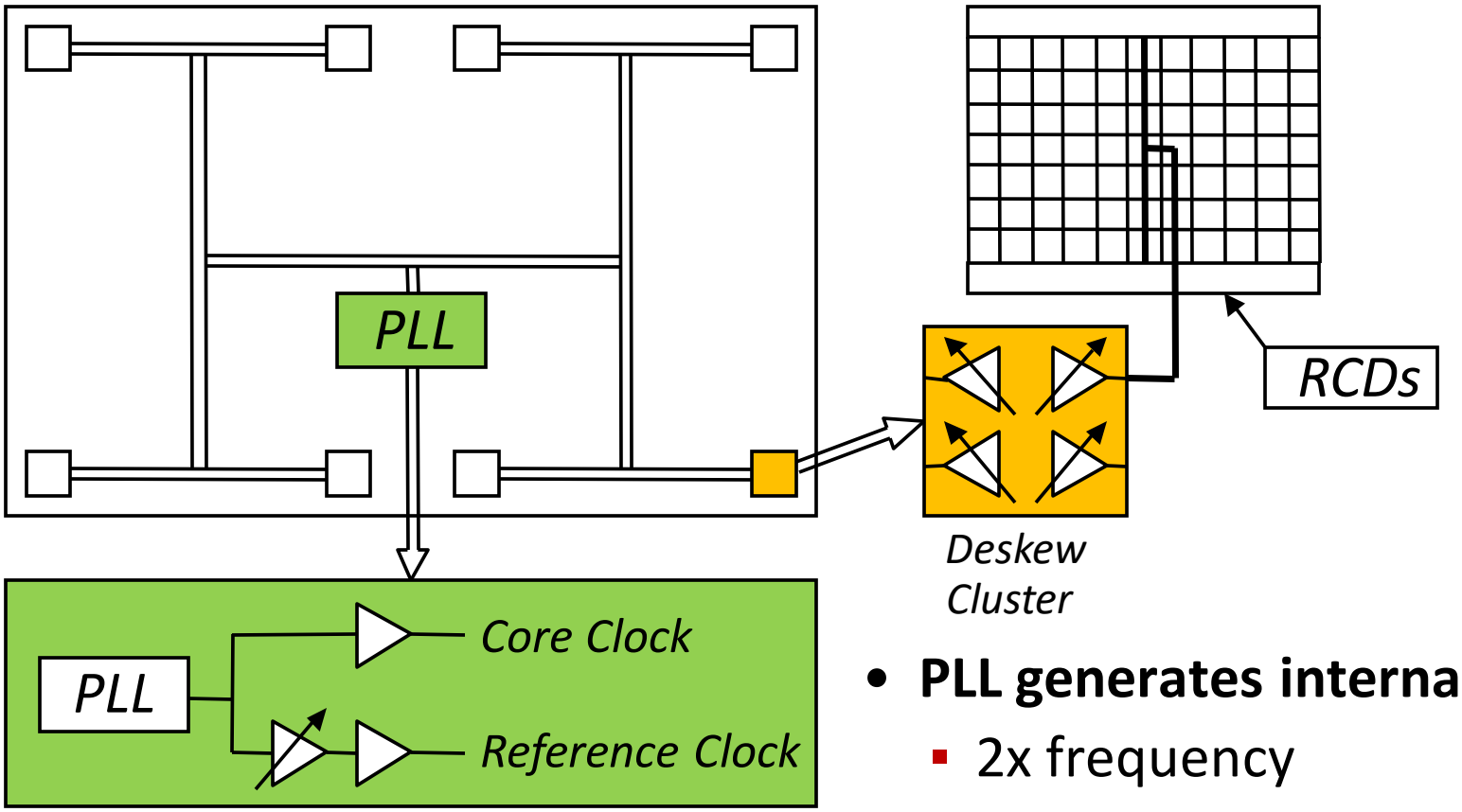
[Geannopoulos and Dai 1998]

IA-32 Pentium Pro: Phase Detector



[Geannopoulos and Dai 1998]

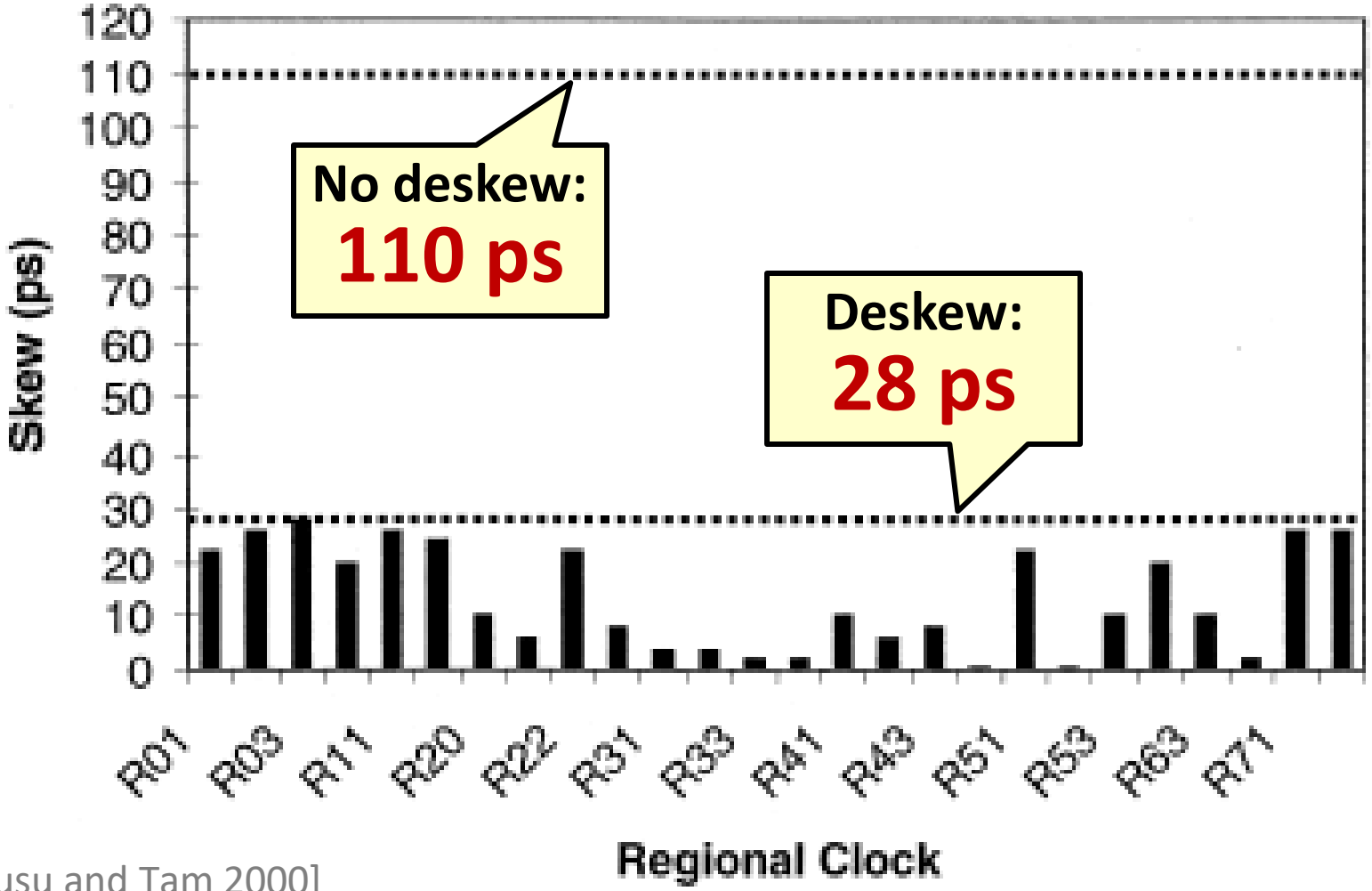
First IA-64 Processor: Clock Distribution



- **PLL generates internal clock**
  - 2x frequency
- **Clk distribution architecture**
  - Balanced global Clk tree
  - Multiple deskew buffers
  - Multiple local Clk buffers

[Rusu and Tam 2000]

Result of Active Deskewing



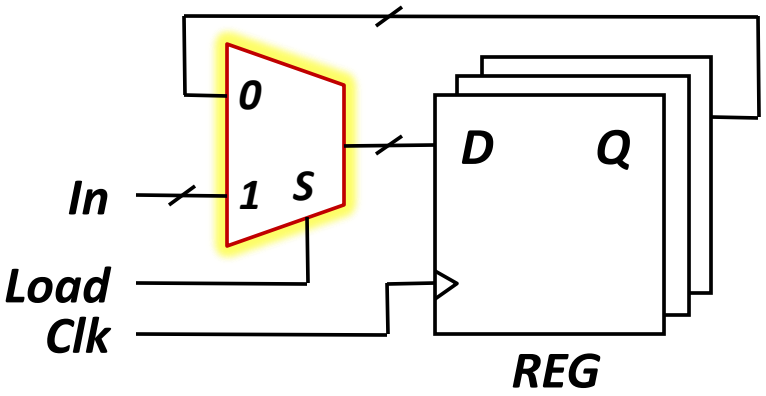
[Rusu and Tam 2000]

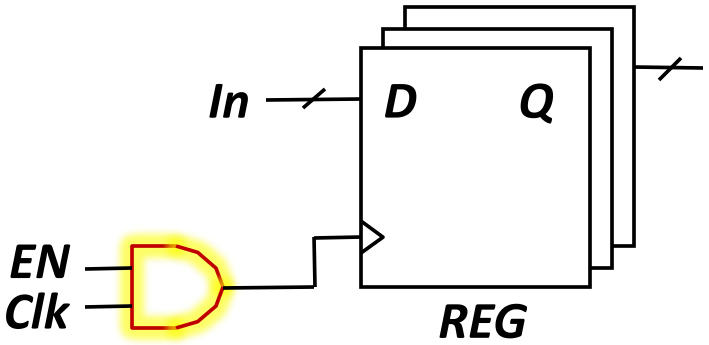
Some Energy Reduction Ideas

- **Clock gating**
  - Global
  - Local
- **Low-swing clocking**
  - Reduced-swing Clk drivers
  - CSE redesign
  - N-only CSE w/ low-swing Clk
- **Dual-edge triggering**
  - Latch-mux
  - Pulsed-latch

Global Clock Gating

Save clocking energy when data activity is low





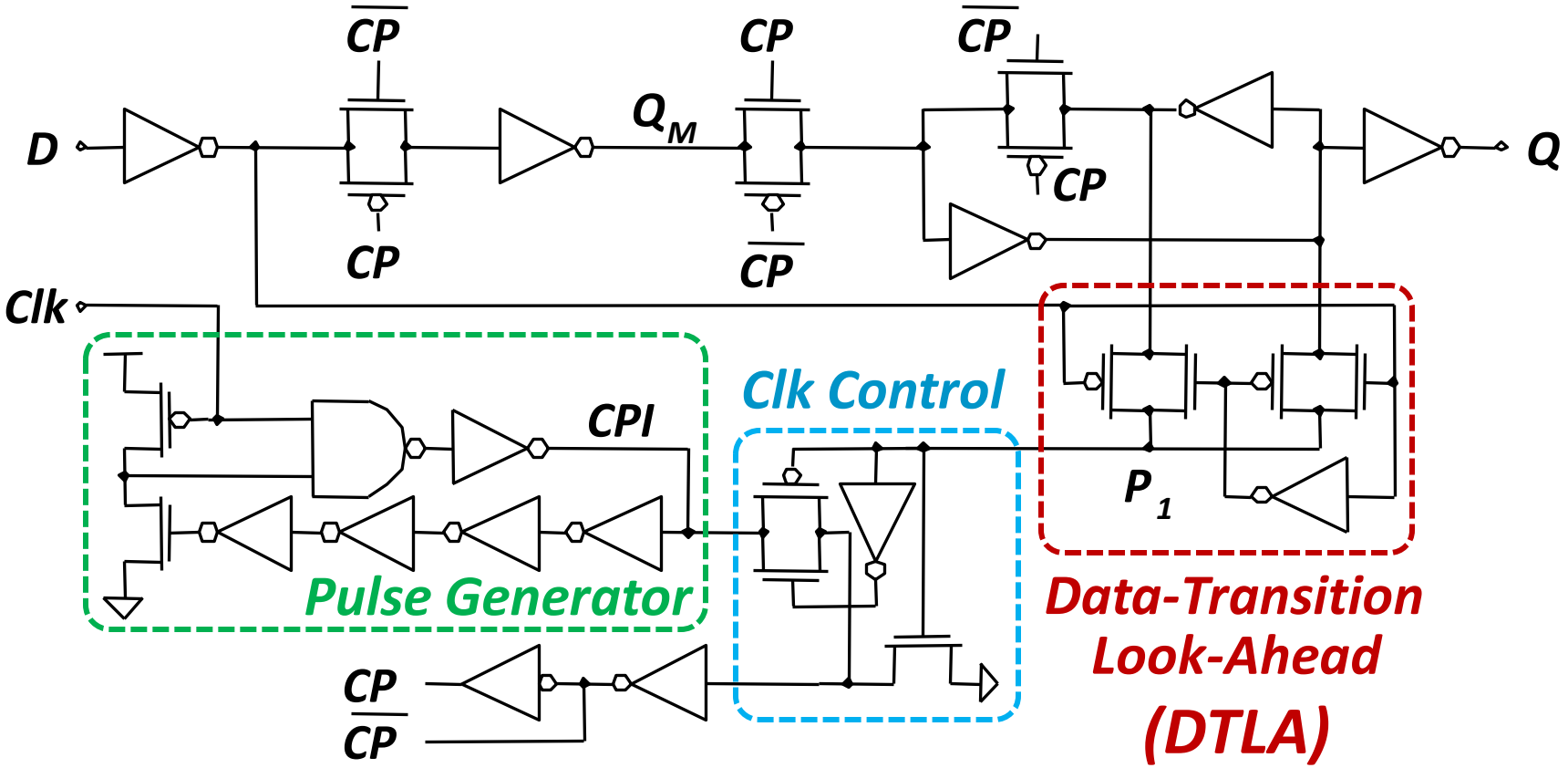
Time-mux
(no gating)

Global Clk Gating

[Kitahara et al. 1998]

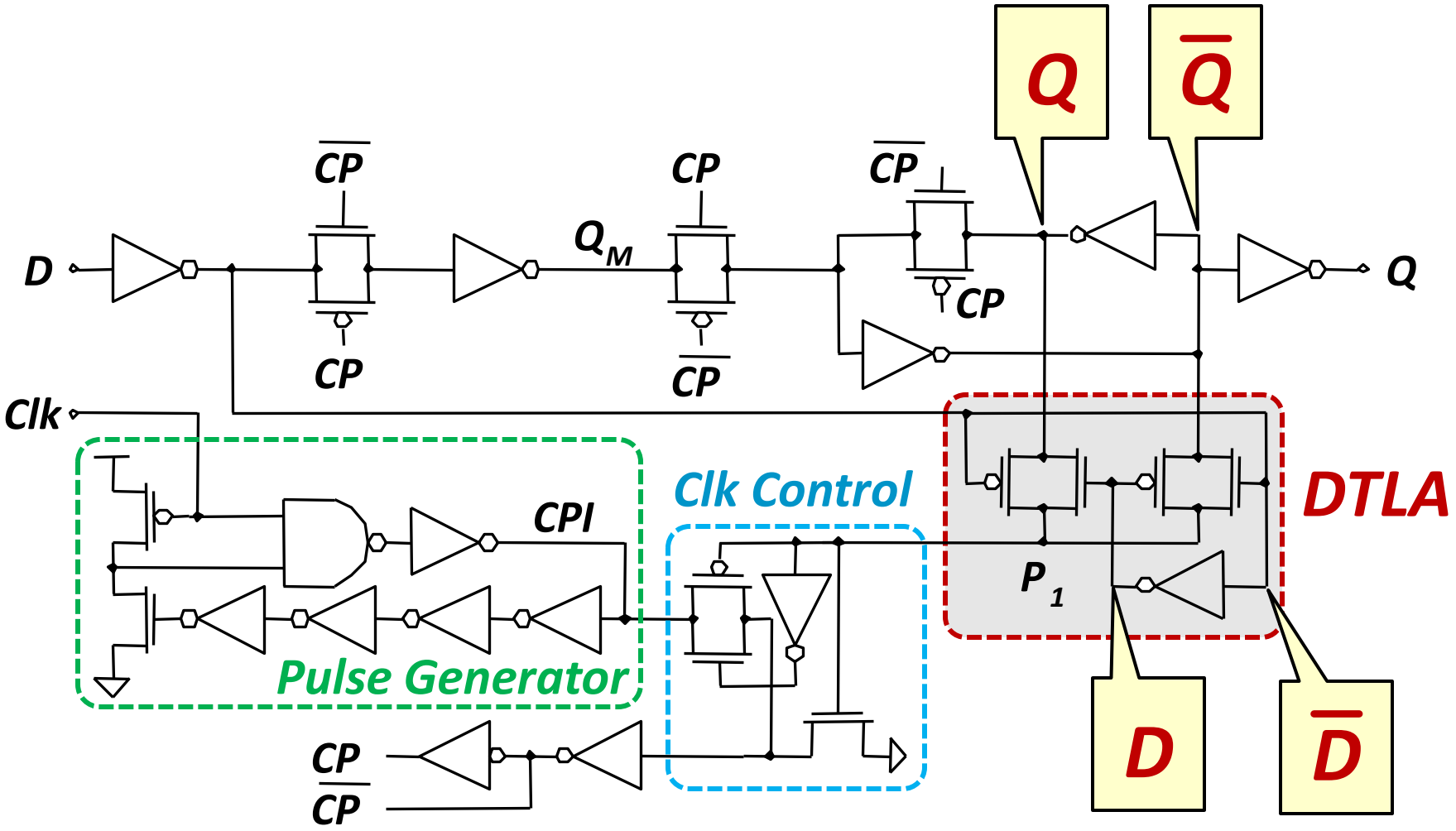
Local Clock Gating (Requires FF Redesign)

Same concept, control at a fine granularity (single FF)



Data-transition look-ahead FF
[Nogawa and Ohtomo, 1998]

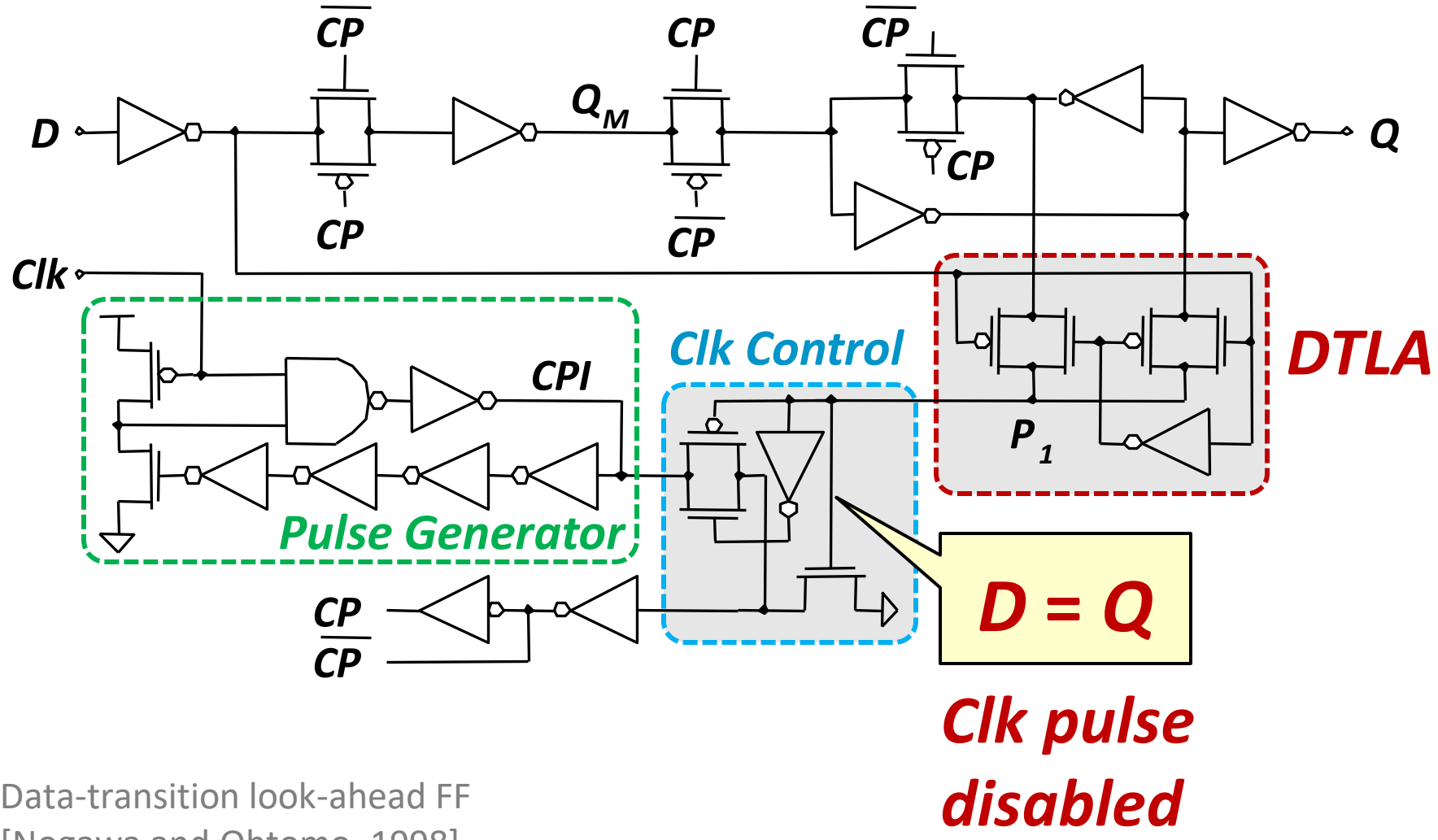
DTLA Circuit: P_1 = D XNOR Q



$$P_1 = \overline{D \oplus Q}$$

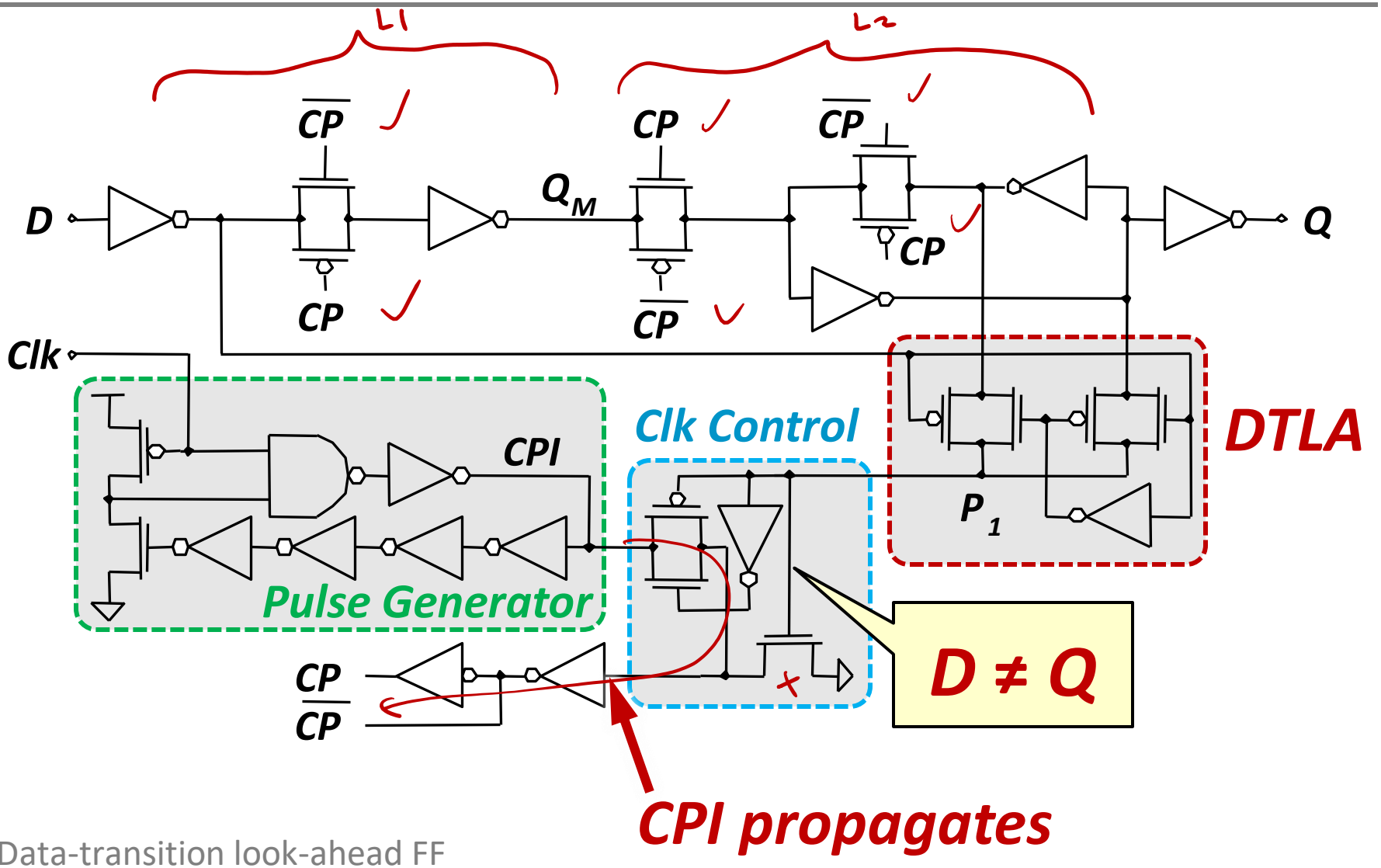
Data-transition look-ahead FF
[Nogawa and Ohtomo, 1998]

Clk Control: Disables Clk When D = Q



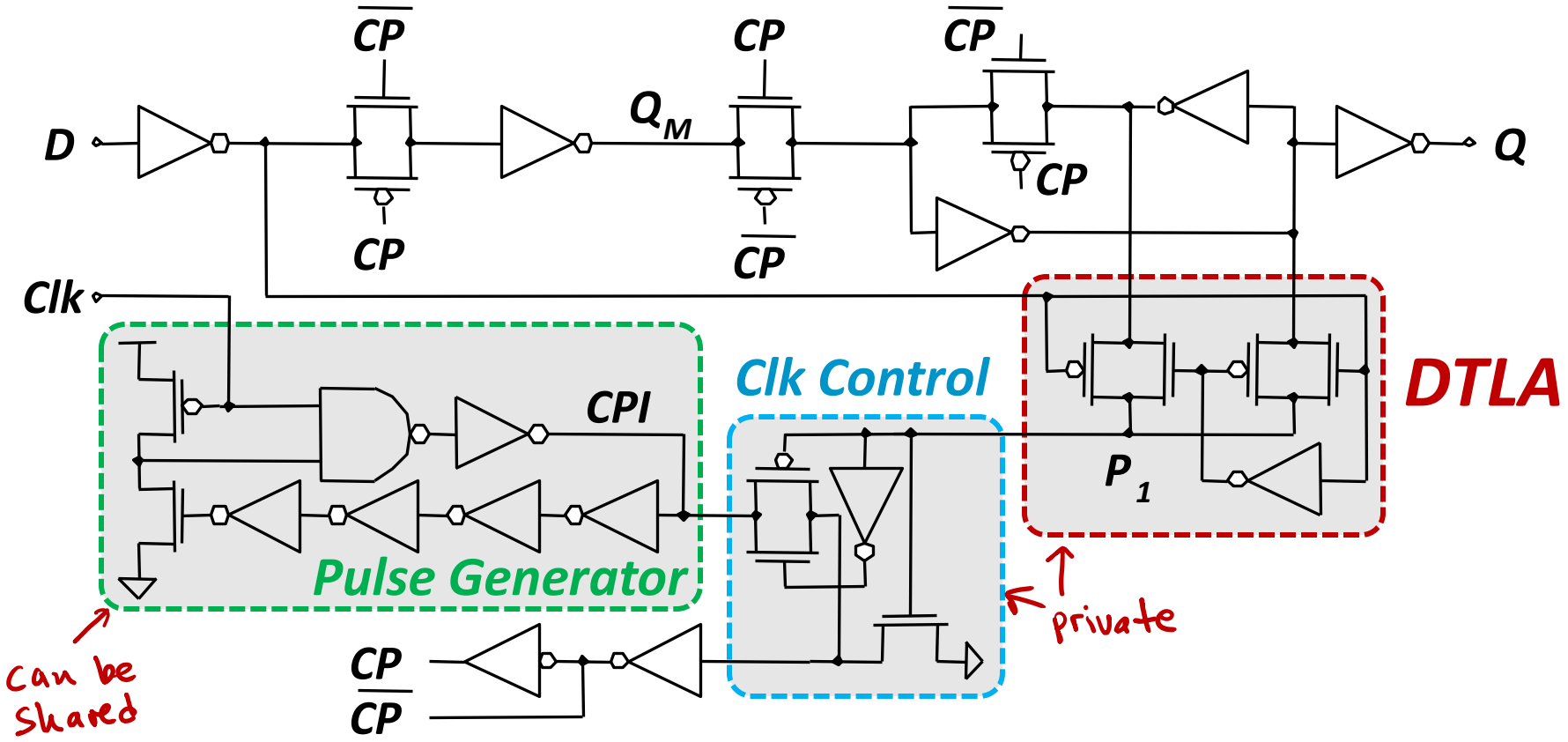
Data-transition look-ahead FF
[Nogawa and Ohtomo, 1998]

Pulse Generator: Pulse at Posedge



Data-transition look-ahead FF
[Nogawa and Ohtomo, 1998]

Data Activity Has to Be LOW



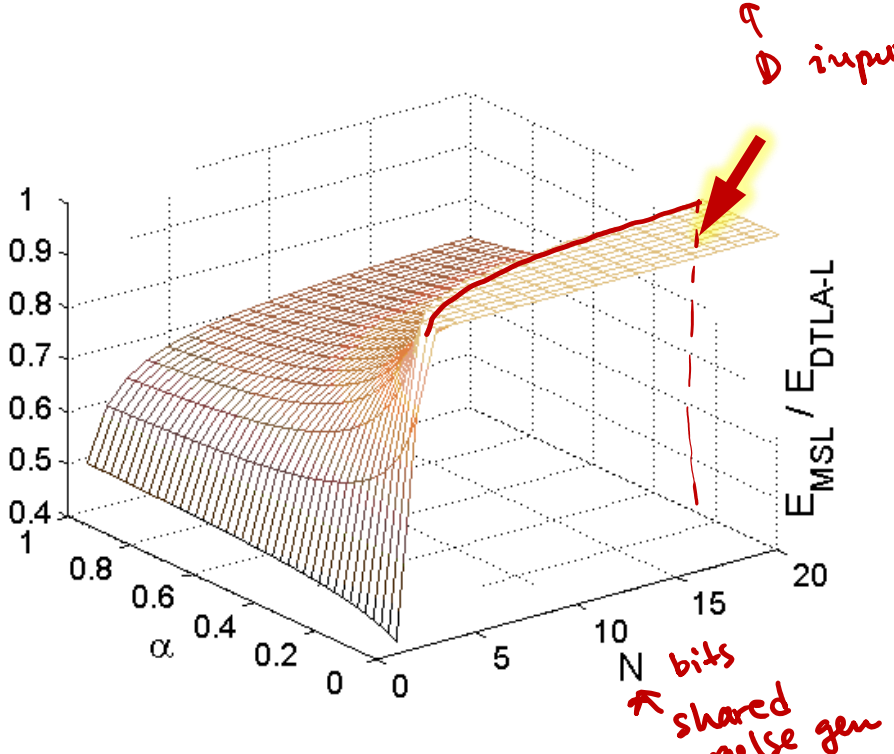
Data-transition look-ahead FF
[Nogawa and Ohtomo, 1998]

Gray circuitry is overhead

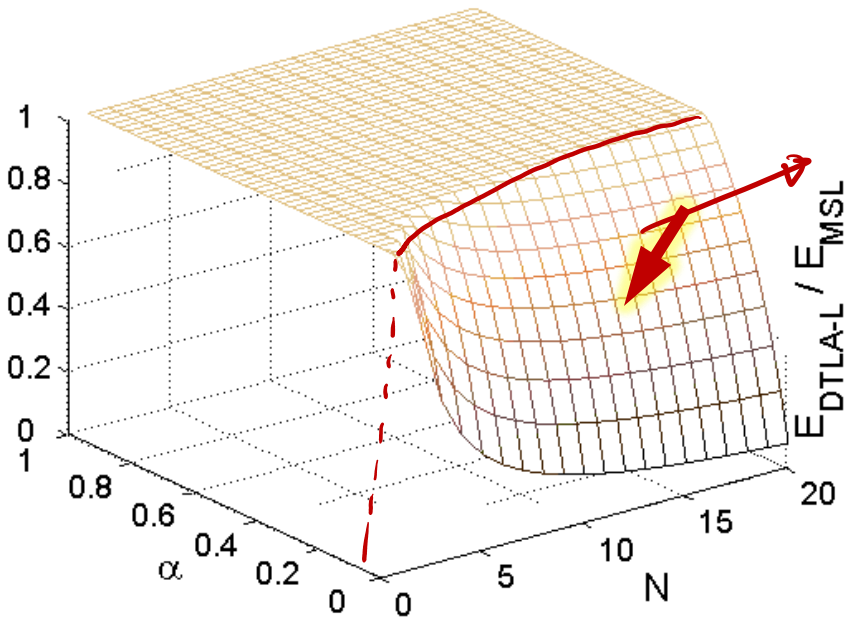
Data activity has to be low

Energy Comparison: DTLA vs. CMSL

- **DTLA-L is more energy-efficient than MS latch-pair (MSL) when N > 2 and α < 0.25**



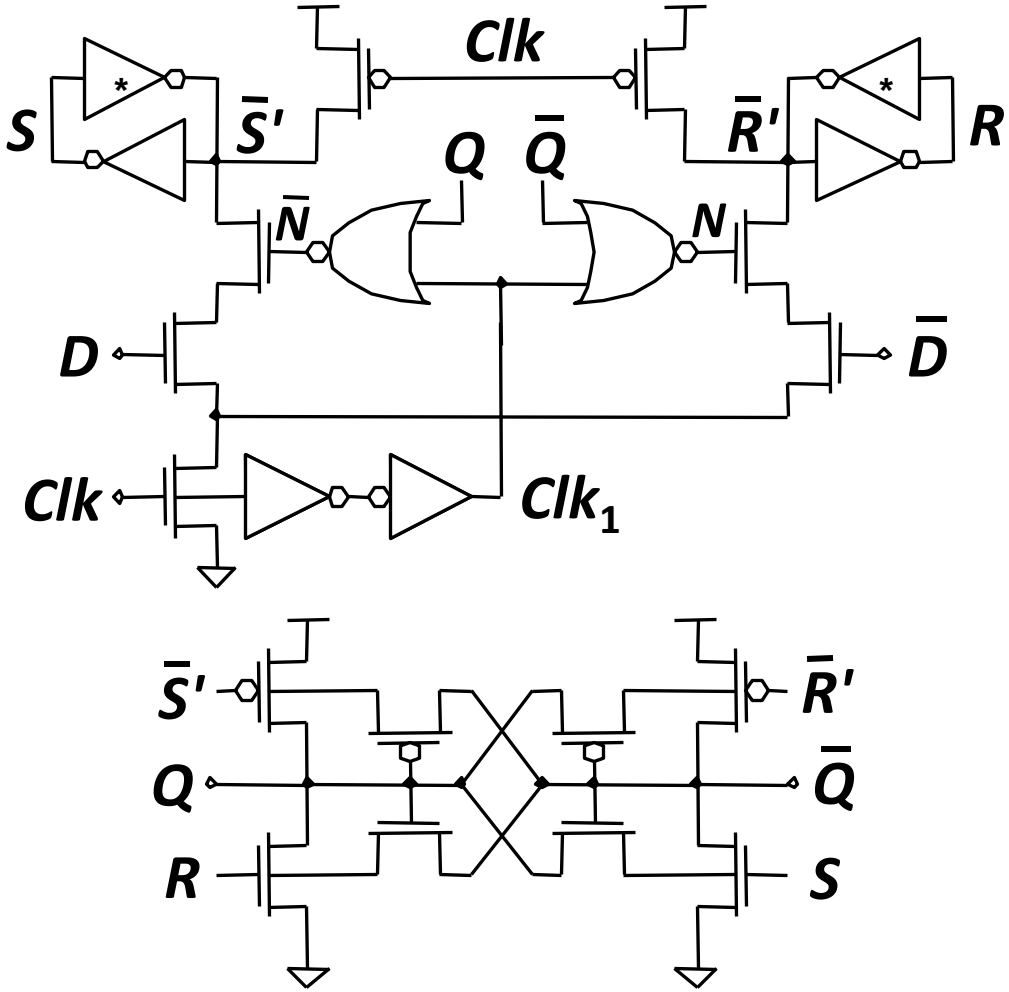
E(MSL) > E(DTLA)



E(MSL) < E(DTLA)

Example 9.5a: Conditional Capture FF

- **Clock enabled only when $D \neq Q$**

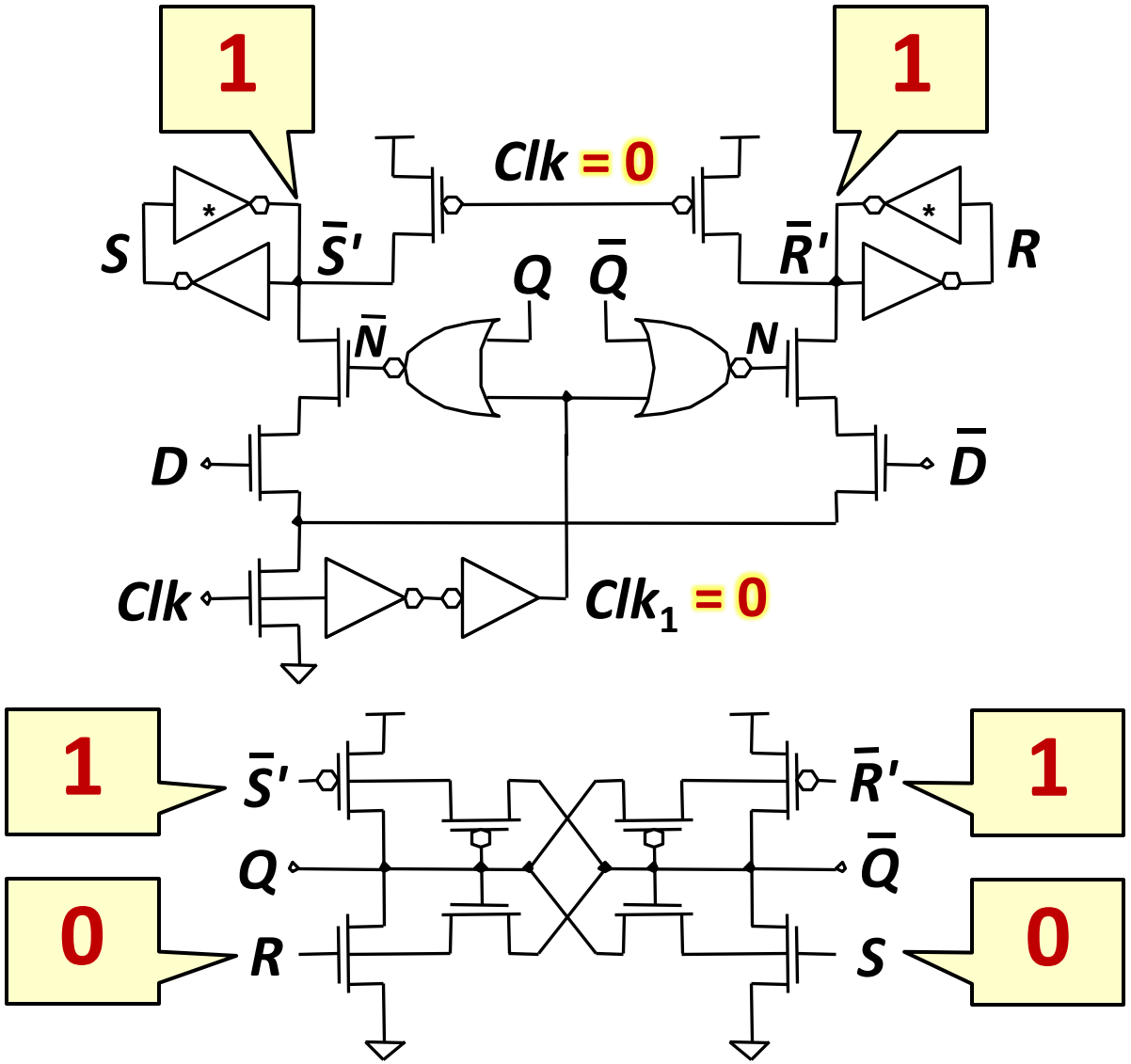


[Kong et al. 2000]

Example 9.5b: Conditional Capture FF

- **Clk = 0**

  Q holds state



[Kong et al. 2000]

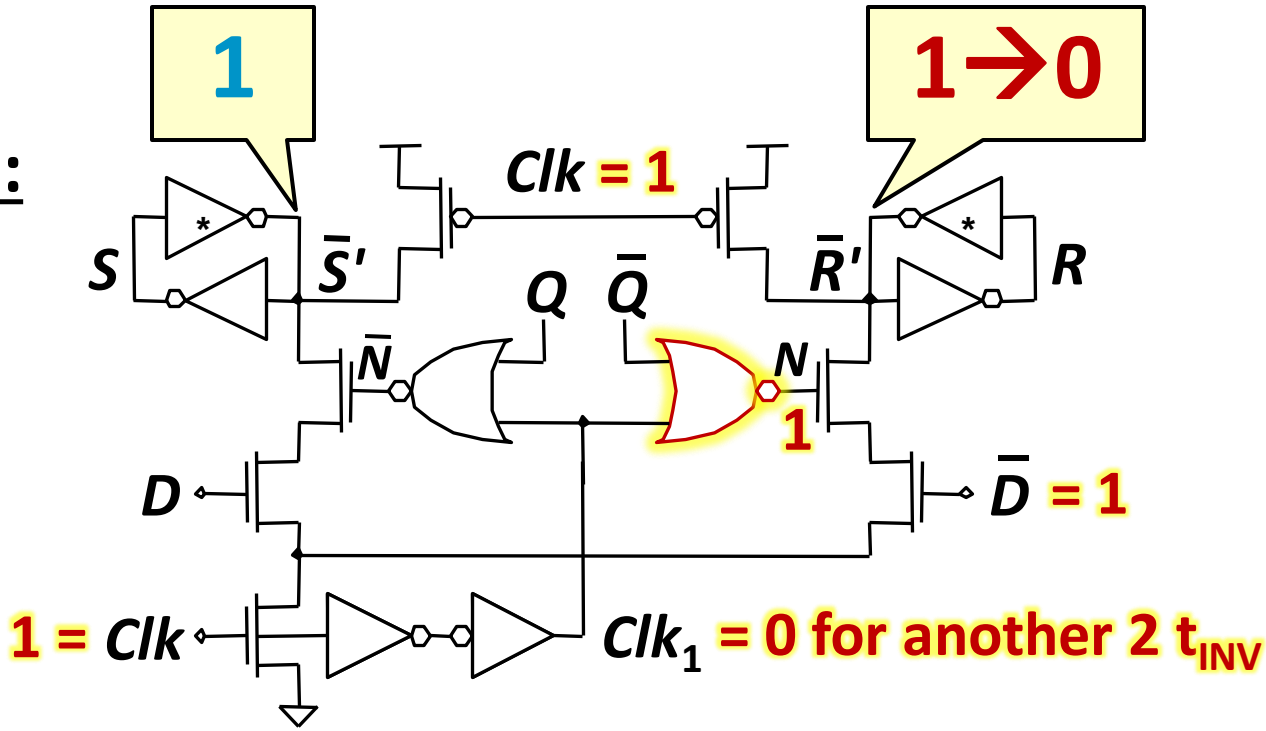
Example 9.5c: Conditional Capture FF

- **Clk = 1**

Example:

Q = 1

D = 0



A branch discharges $\bar{S}$' or $\bar{R}$' if $D \neq Q$

(Clk is propagated into the FF)

[Kong et al. 2000]

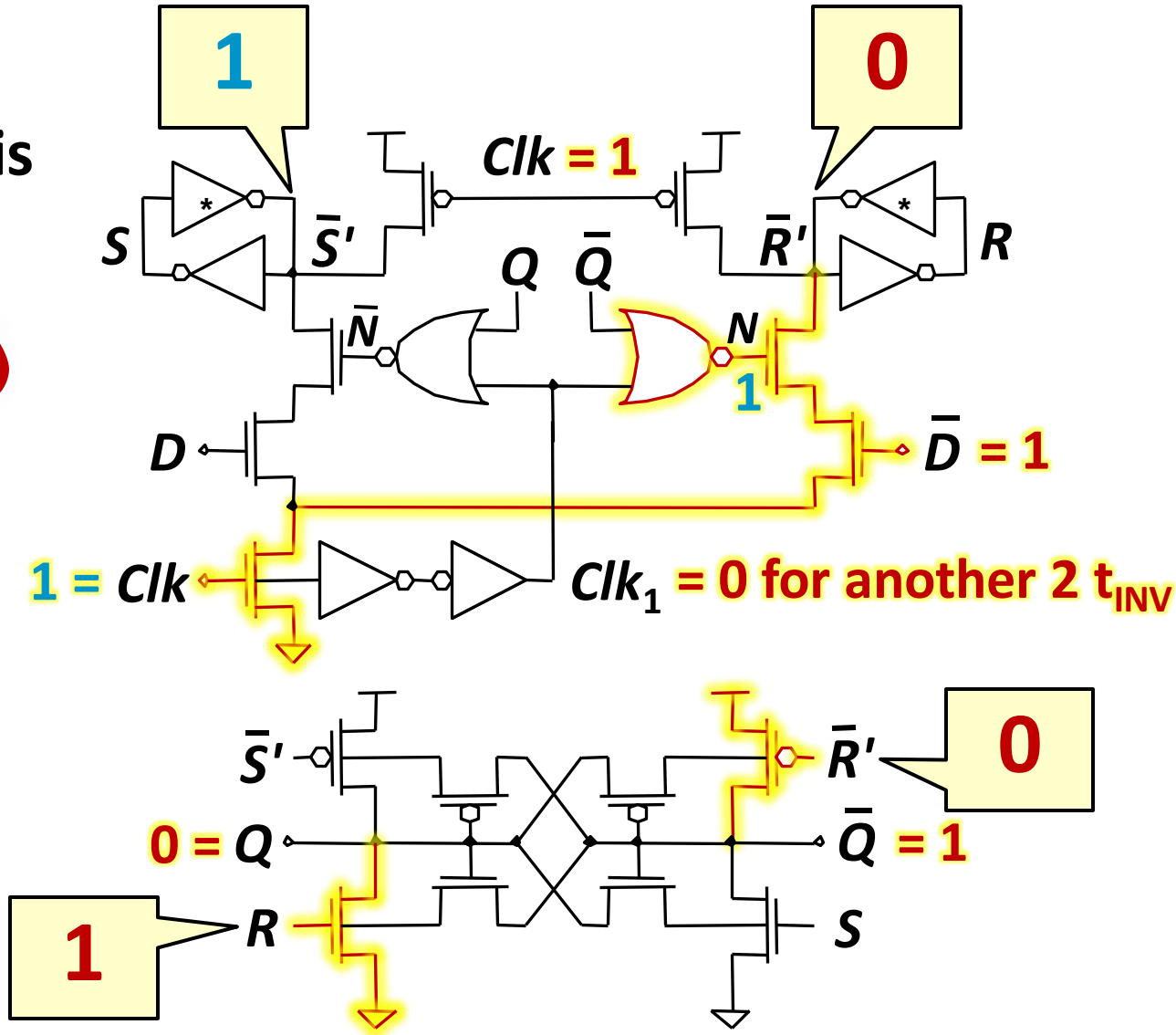
Example 9.5d: Conditional Capture FF

- **Clk = 1**

  New Q is stored

  Q = 0

  D = 0

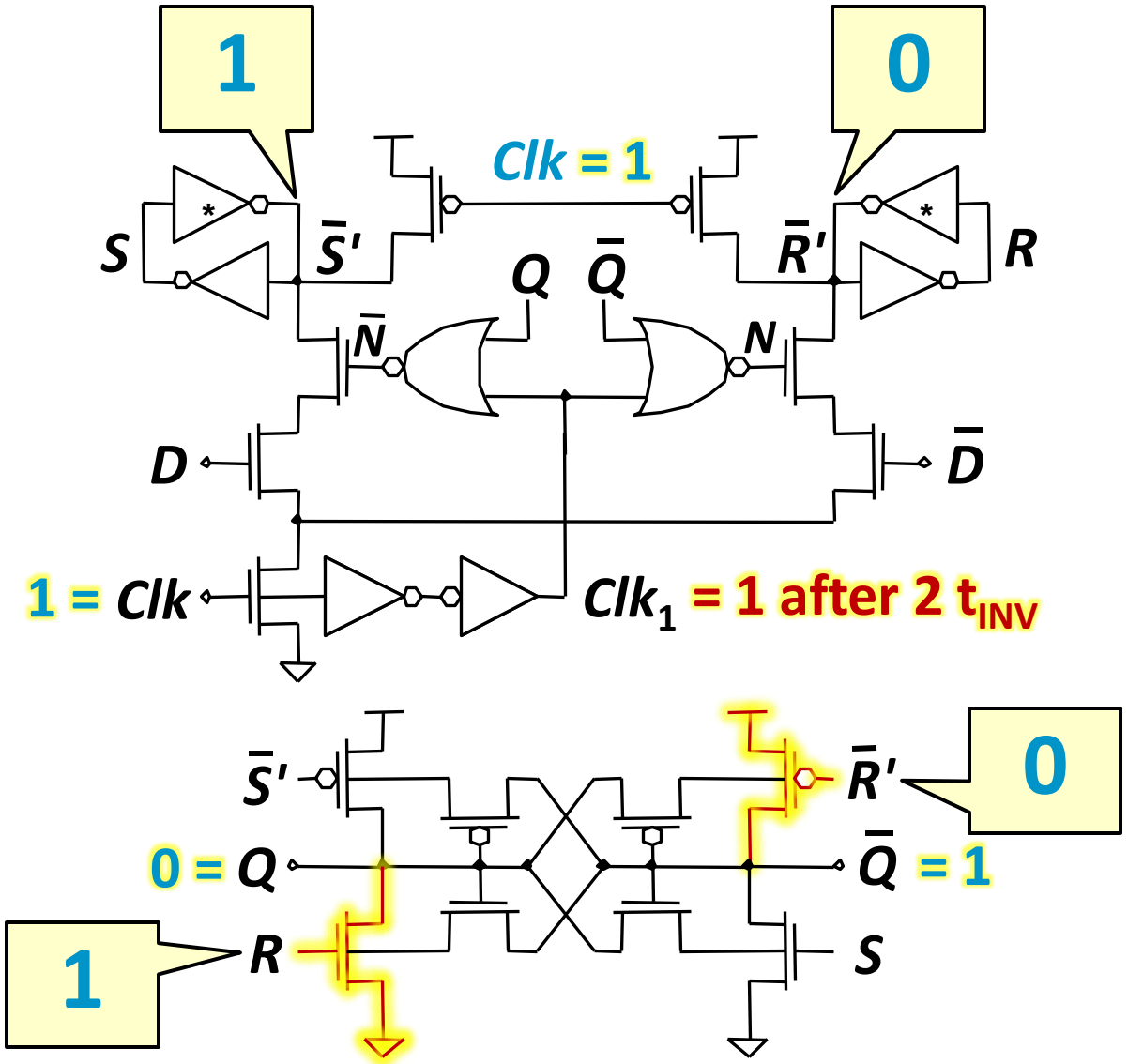


[Kong et al. 2000]

Example 9.5e: Conditional Capture FF

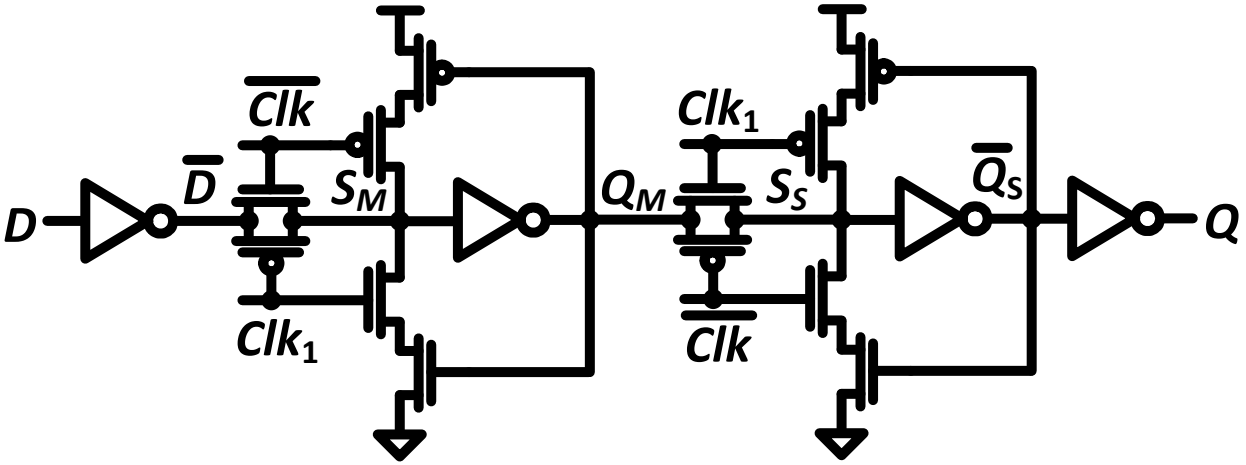
- Clk = 1

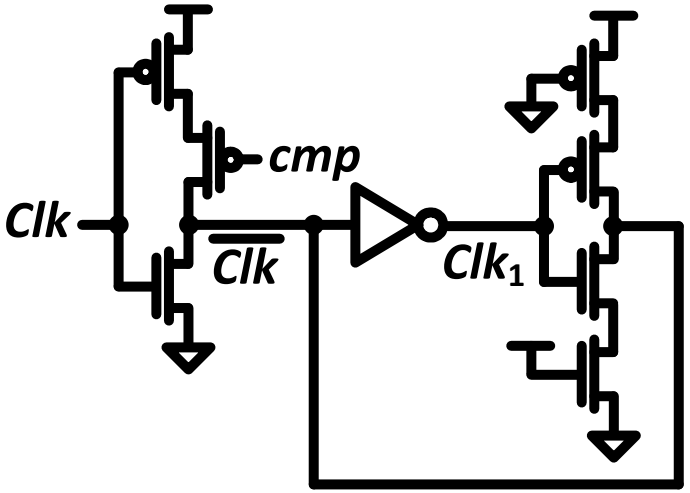
  Q = 0

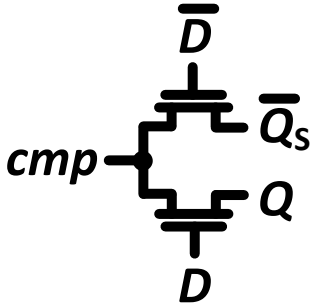


[Kong et al. 2000]

Gated Master-Slave Flip-Flop



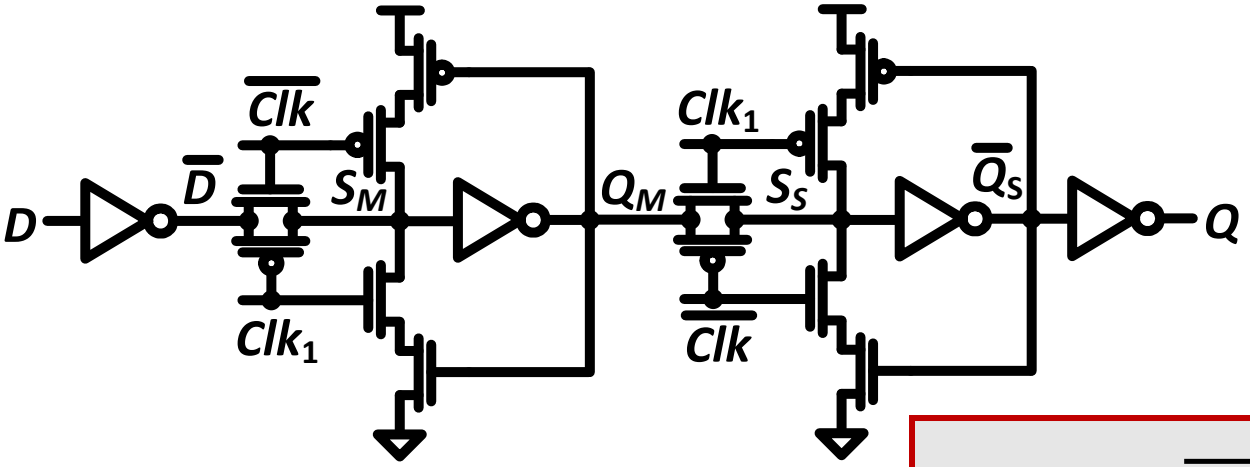




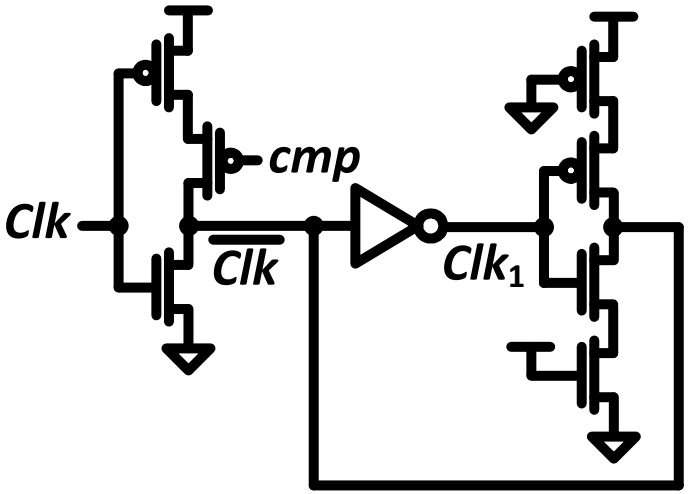
[Marković et al. 2001]

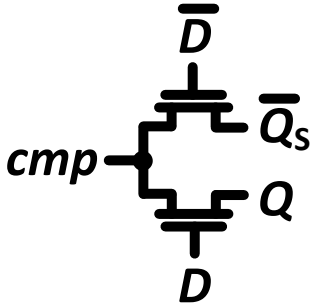
Gated Master-Slave Flip-Flop

- **D = Q : cmp = 1**, $\overline{Clk} = 0$, $Clk_1 = 1$ (M closed, no Q change)



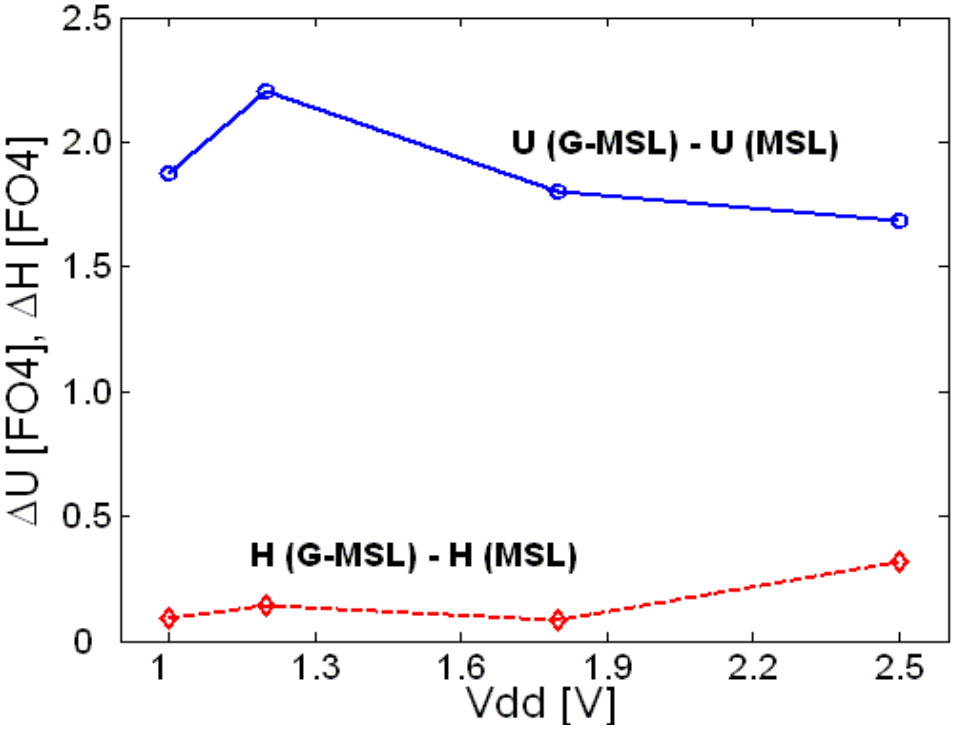
$$cmp = \overline{D \oplus Q}$$





[Marković et al. 2001]

Gated TG MS Latch: Timing

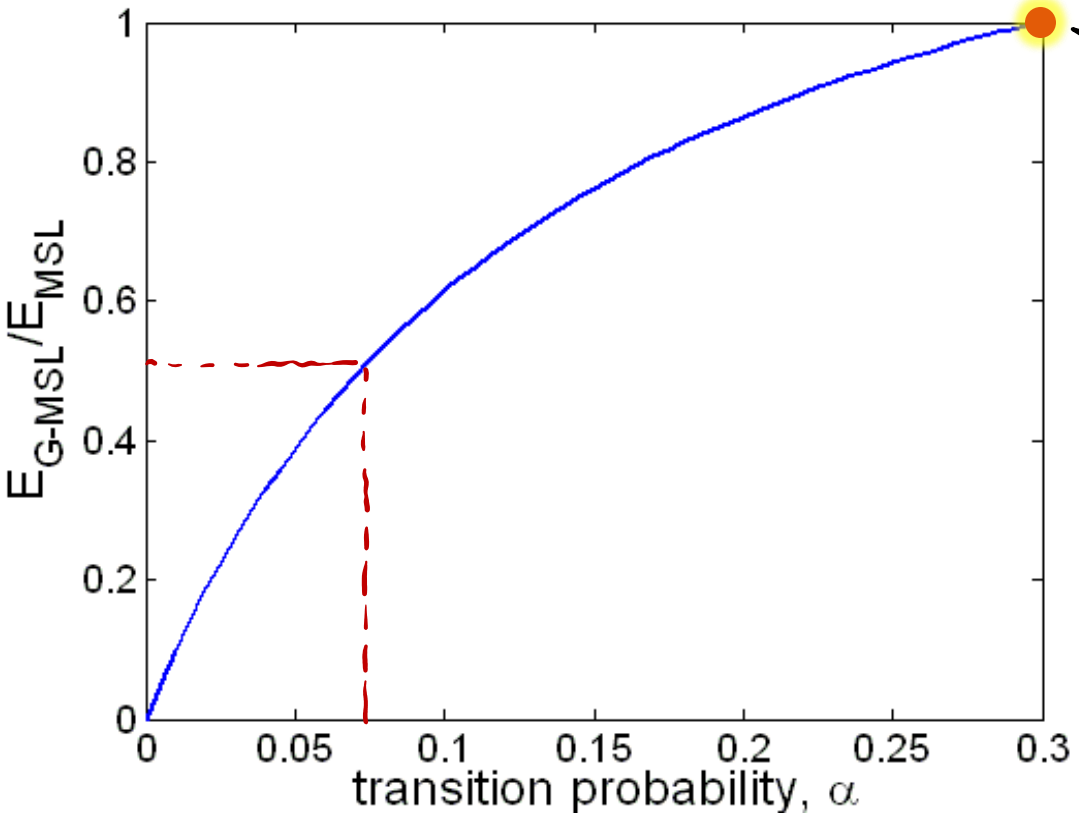


Setup (U)

Hold (H)

- **Increased t_{setup} in gated MSL**
  - Comparator inserted into the critical path

Gated TG MS Latch: Energy

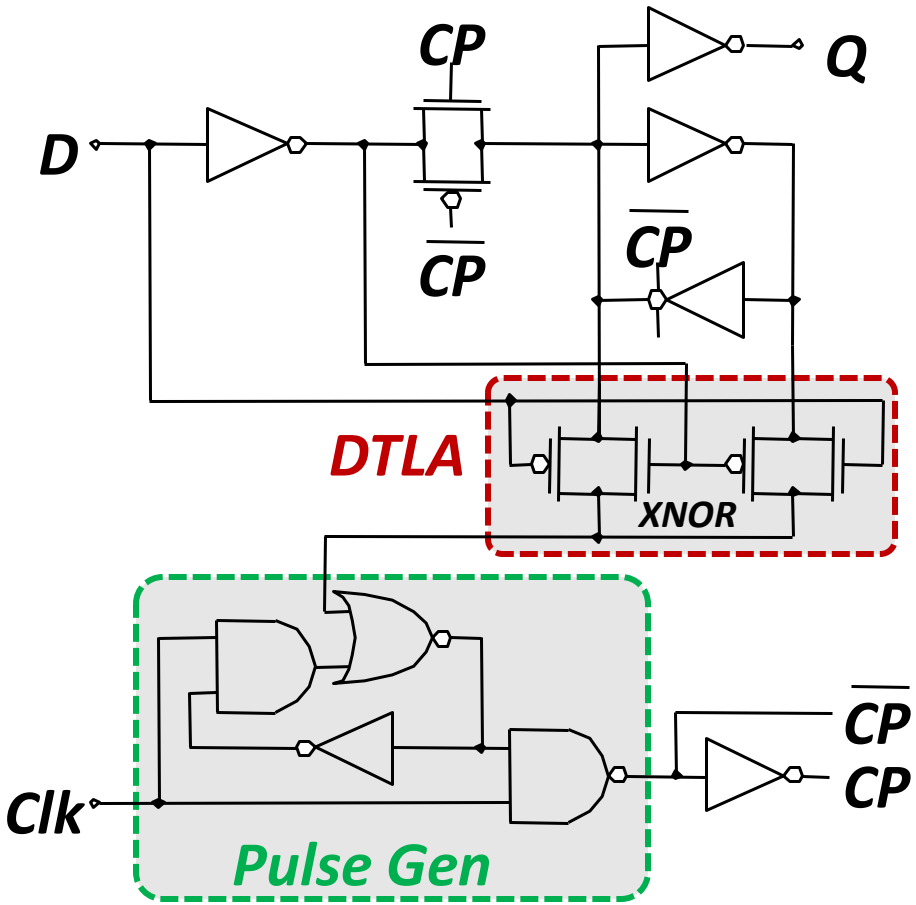


Lower Energy for $\alpha < 0.3$

Otherwise cmp and Clk gen dominate energy

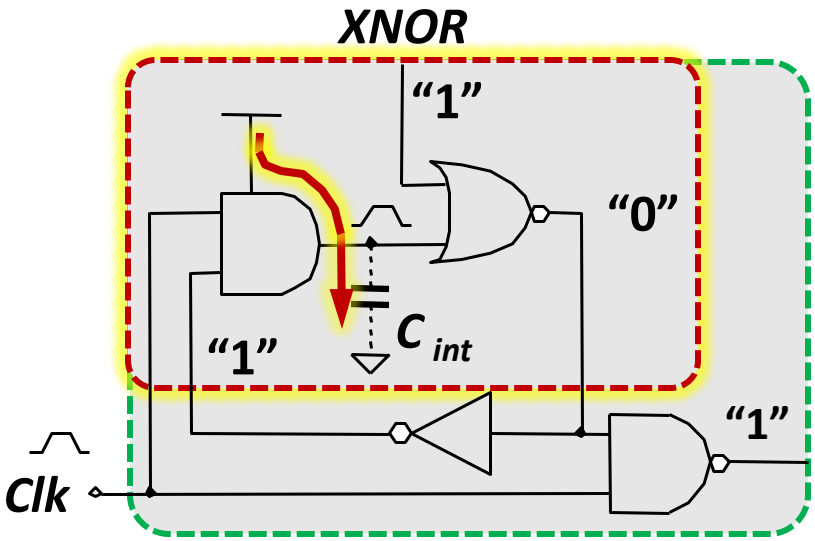
Clock-on-Demand (COD) Pulsed-Latch

- **Pulsed latch, Clk pulses gated with XNOR ckt**
  - **D ≠ Q ⇒ XNOR = 0**, CP→1 when Clk↑, and CP→0 after Q has changed to D
  - **D = Q ⇒ XNOR = 1** ⇒ CP = 0, no Q change or energy consumption
- **Pulse Generator includes Clk control**
  - Cannot be shared

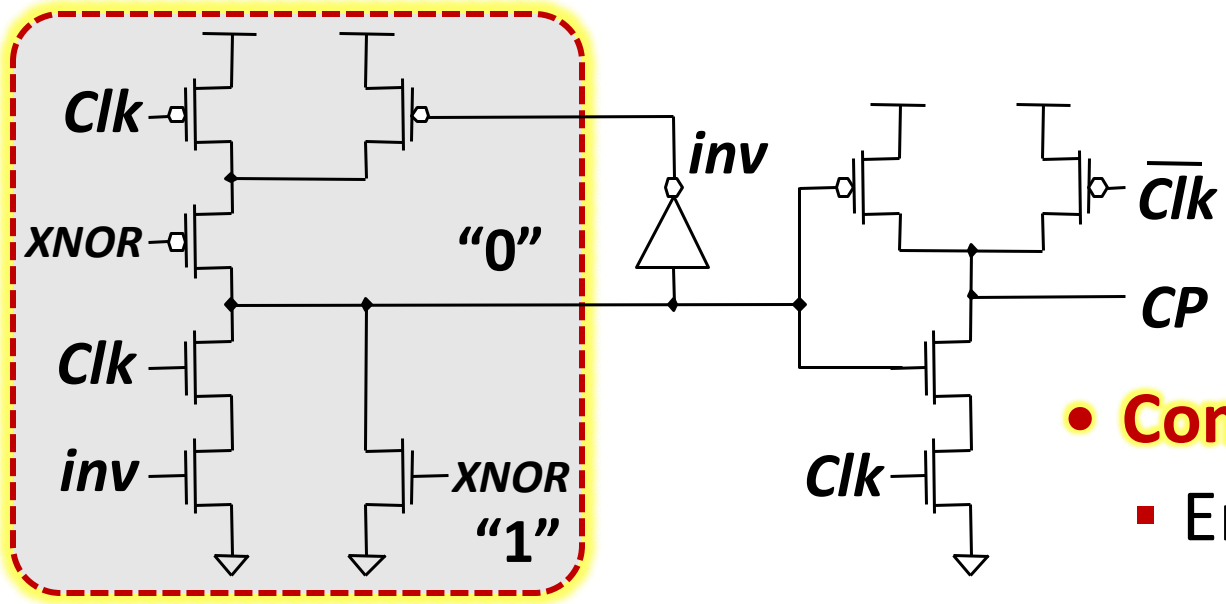


[Hamada et al., 1999]

Energy-Efficient Pulse Generator in COD-PL



- **Straightforward realization (CMOS gates)**
  - C_{int} switches every cycle
  - Energy-inefficient

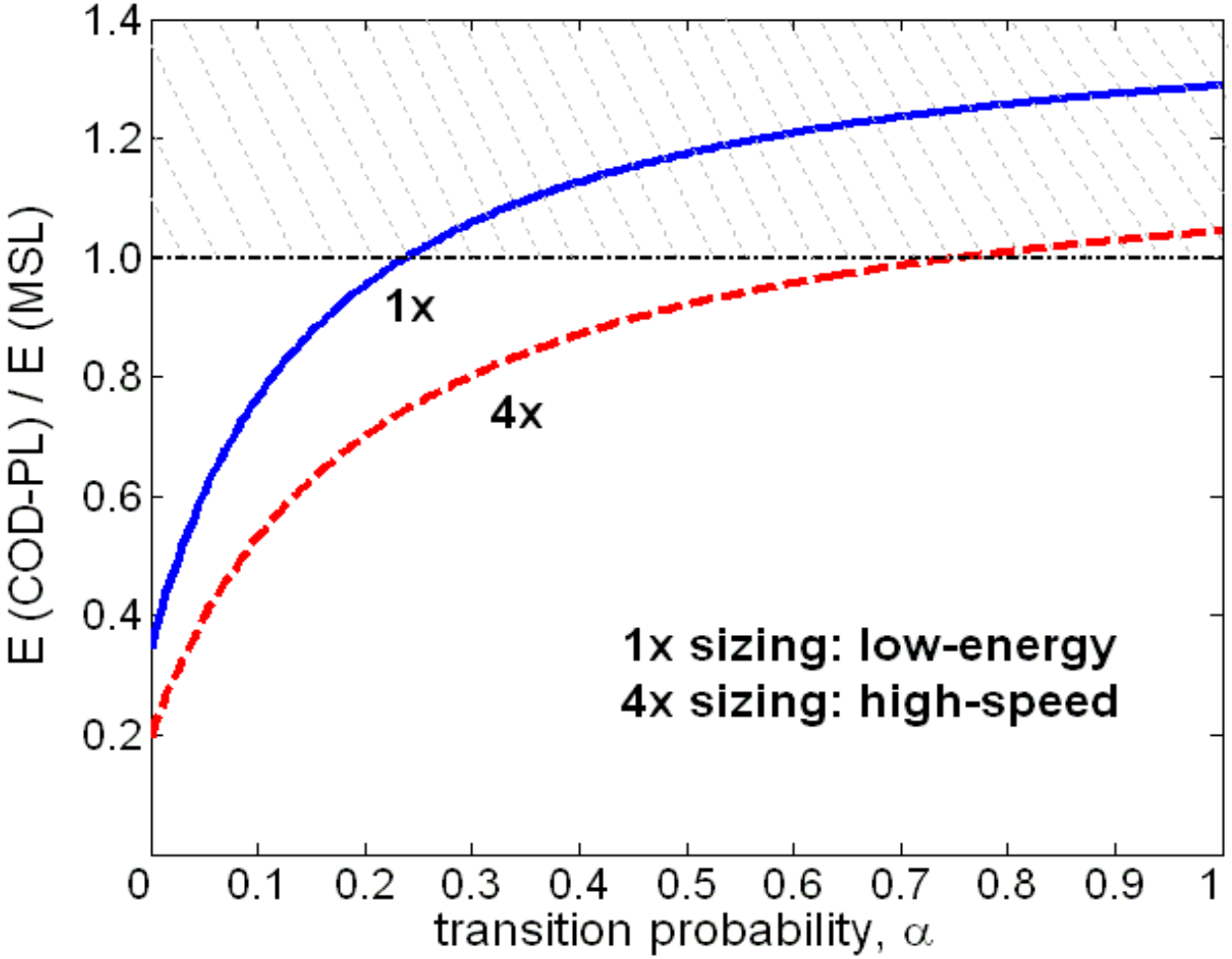


- **Compound AND-NOR**
  - Energy-efficient

Not Just Switching Activity:
Output Loading Matters

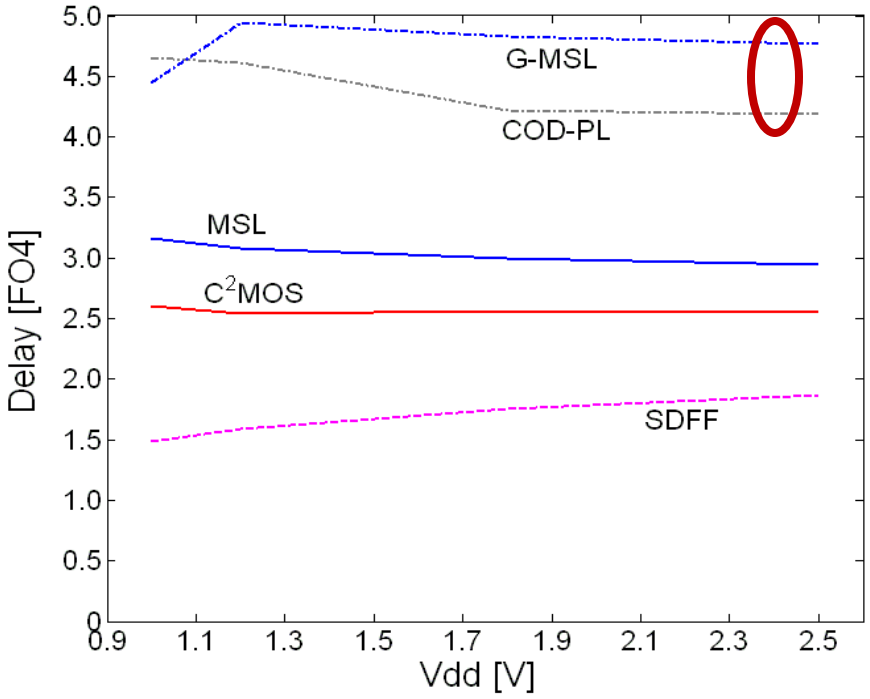
Sizing Impacts Energy Efficiency

Sizing for high-speed (4x): lower impact of Clk transistors



[Marković et al., 2001]

Timing Comparison: Delay

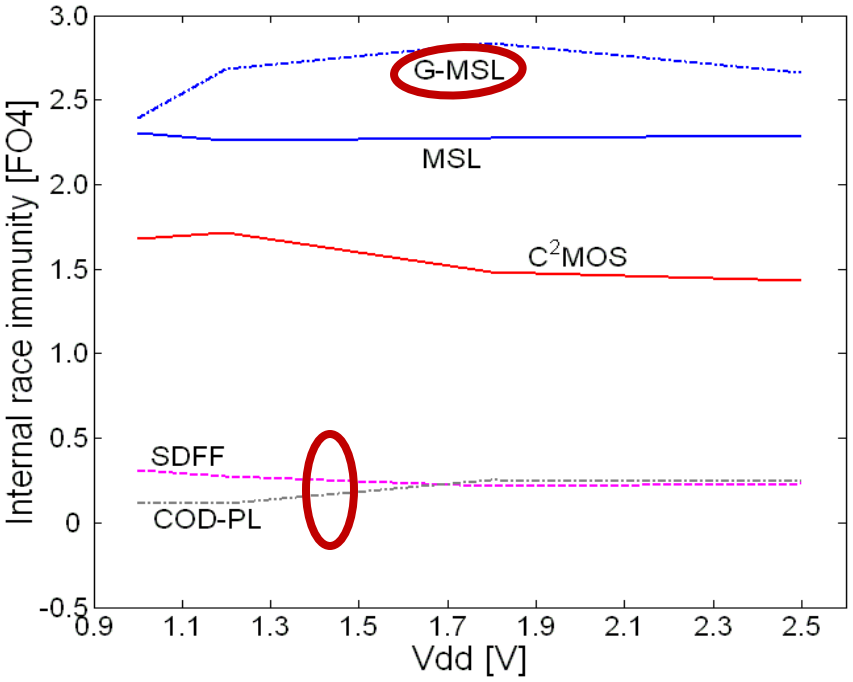


Clk gating designs:

- **Delay constant w/ V_{DD}**
- **Large D due to long t_{setup}**

[Marković et al., 2001]

Timing Comparison: Delay



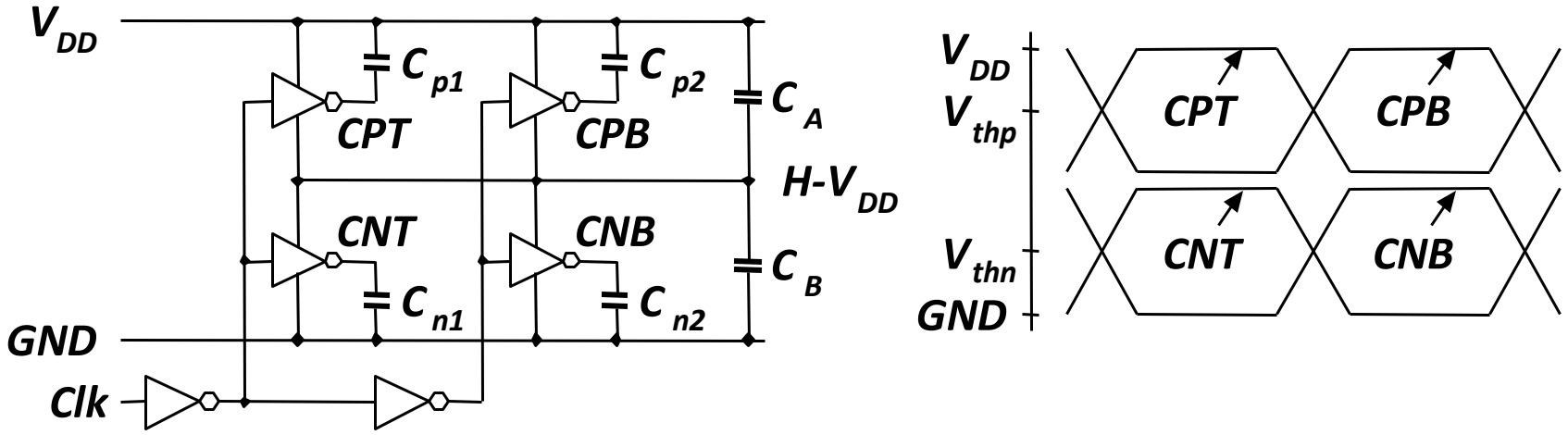
Race immunity:

- **PL < MS < Gated MS**
- **CLD-PL bad (wide Clk pulse)**

[Marković et al., 2001]

Low-Swing Clocking: Driver Re-design

- **Half-swing clock drivers:** **50% power reduction (minus some penalty in clock drivers)**



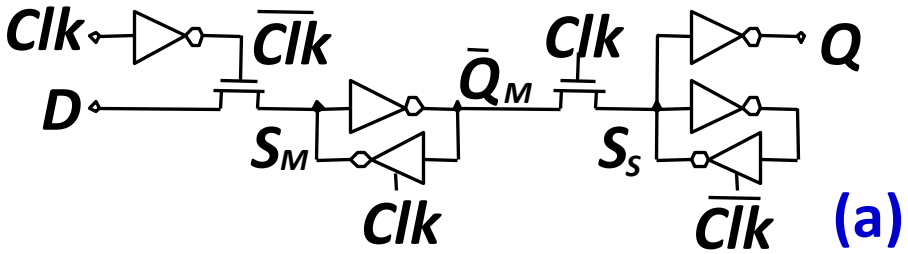
[Kojima at al. 1995]

N-only Clocked Latches: Idea

- **NMOS transistors for reduced clock swing**
  - Avoid conflict with partially turned-off PMOS
- **Reduced clocking energy, some penalty in delay**
  - *Clk* is in the critical path

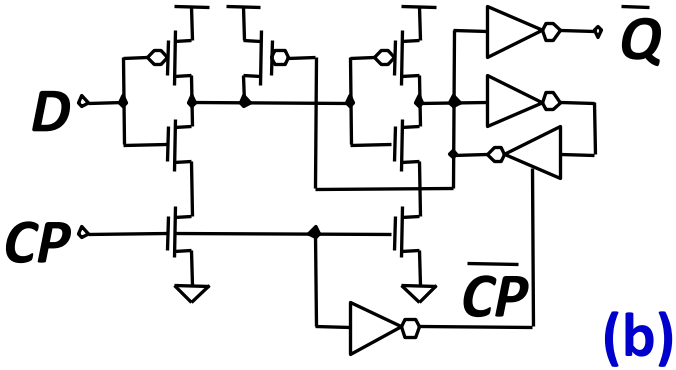
N-only Clocked Latches: Circuit Examples

Conventional TG MSL (N-MSL)



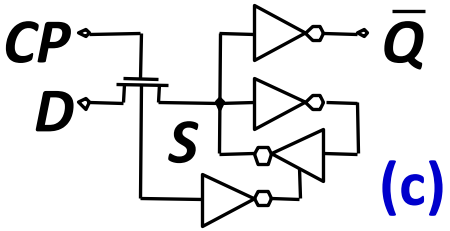
(a)

Pulsed-latch (N-FF)

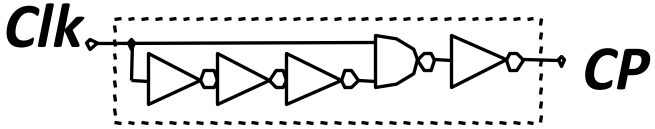


(b)

Conv. PL (N-PL)

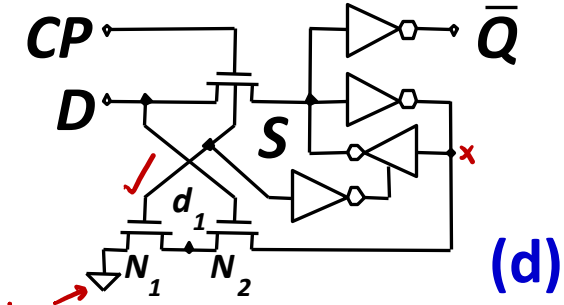


(c)

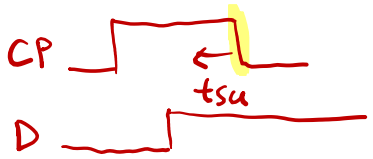


Pulse Gen (b)-(d)

Push-Pull PL (N-PPL)



(d)



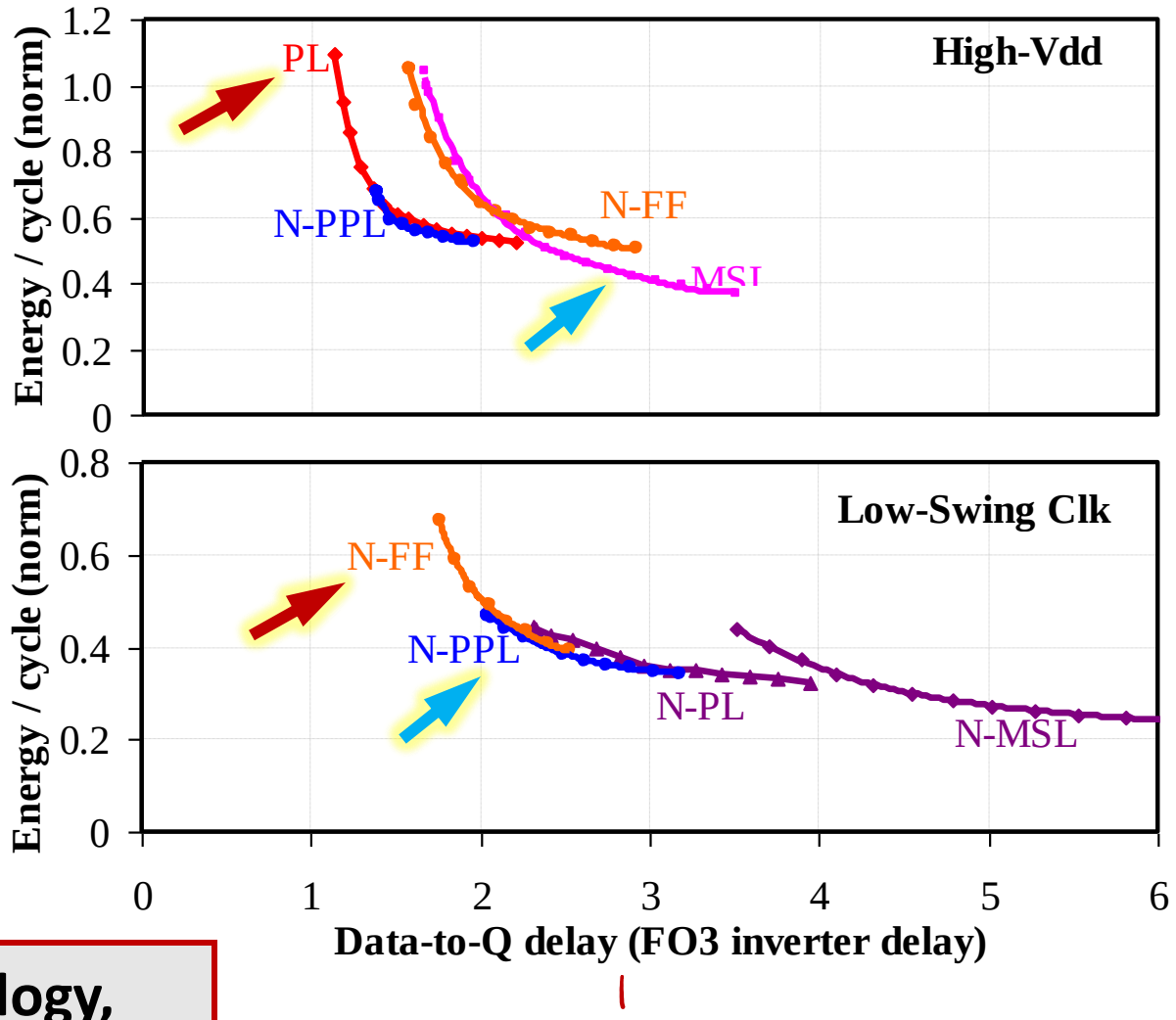
Low-Clk-Swing FFs: Energy-Delay Comparison

Full-swing:

- **PL** preferred for **high-speed**
- **MSL** preferred for **low energy**

Low-swing clock:

- **N-FF** preferred for **high-speed**
- **N-PPL** preferred for **low energy**

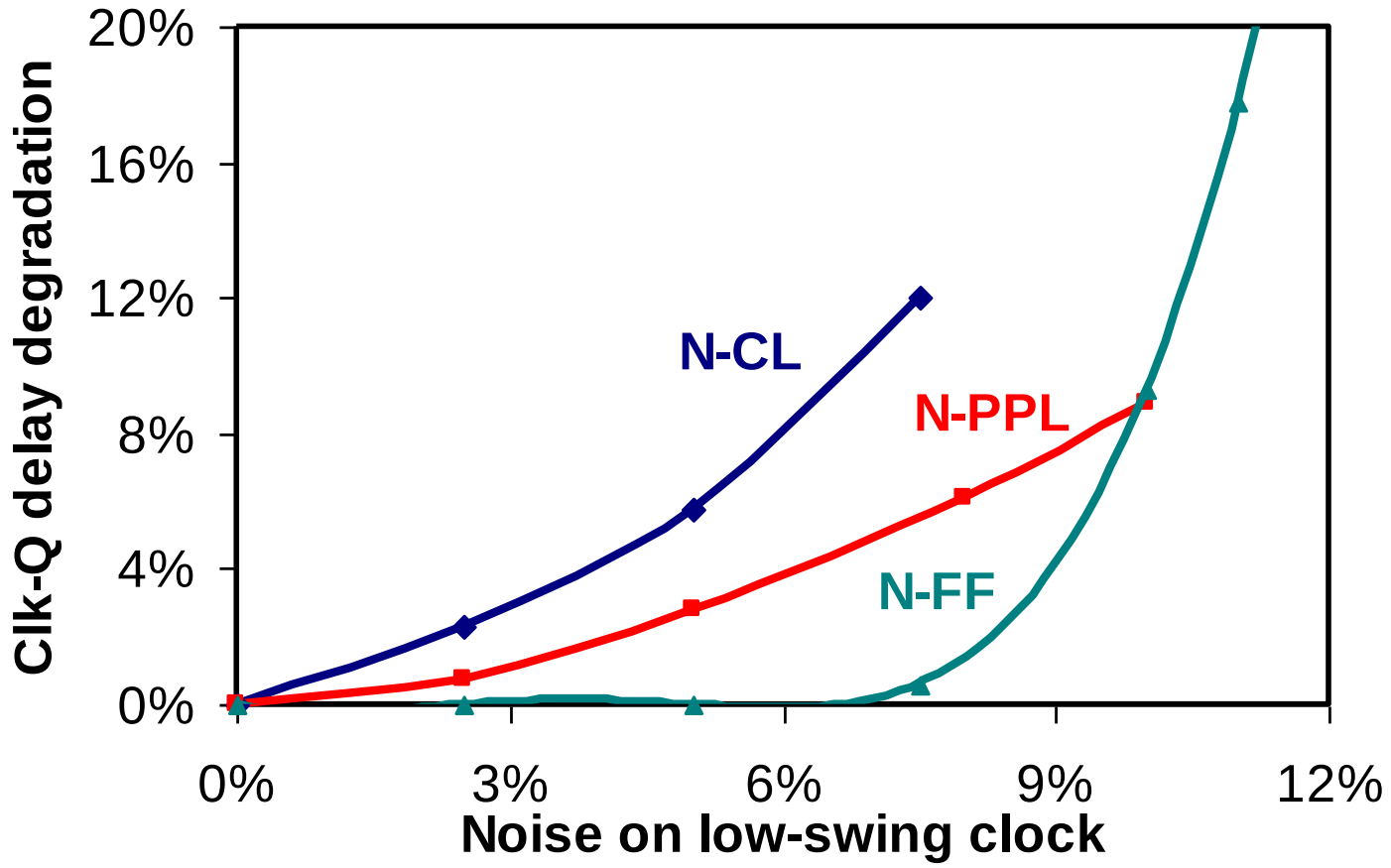


[Marković et al. 2004]

Data: 130nm technology,
C_L = 50fF, $C_{in} \leq$ 12.5fF, α = 0.1

Effect of Clock Noise on Latch Delay

- All FFs fail for clock noise > 12% of clock voltage
- N-FF gives best clock noise rejection

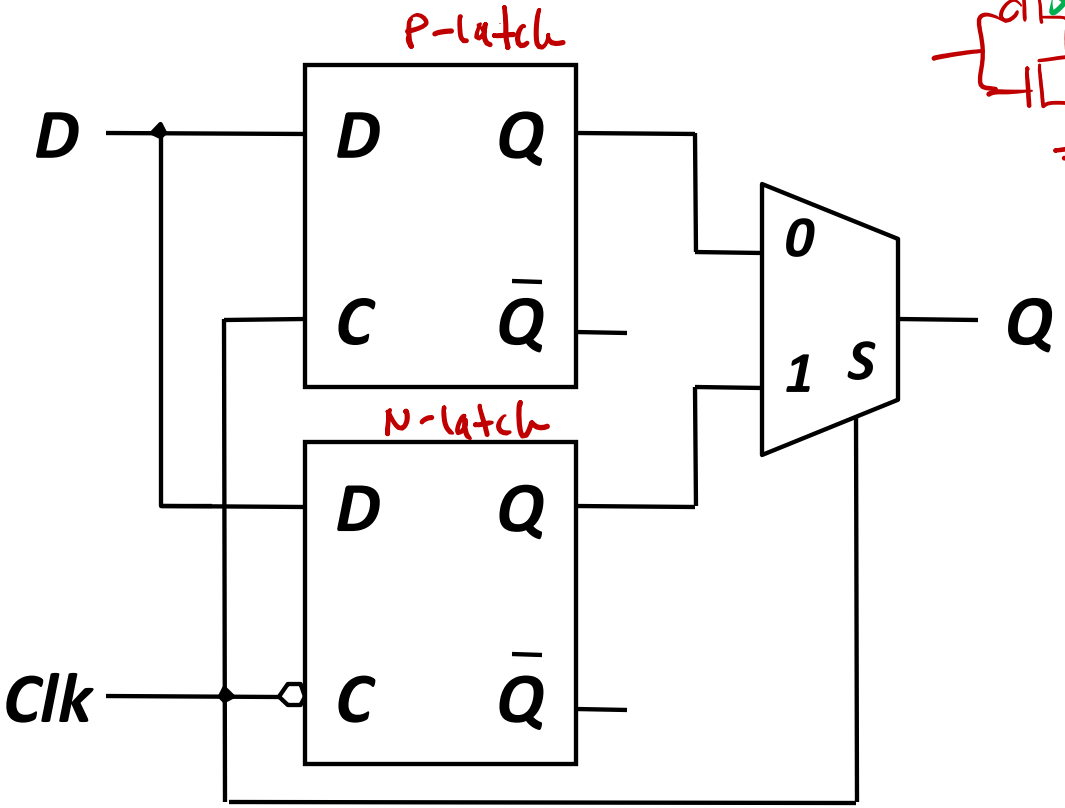


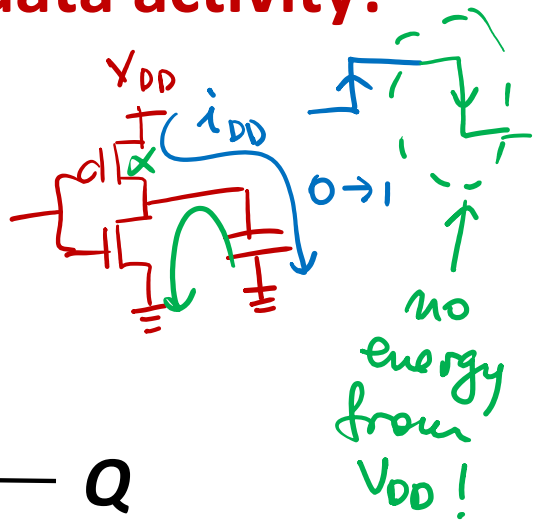
Dual-Edge Triggering

- Latch-Mux
- C^2MOS Latch-Mux
- Pulsed Latch
- SR-based Flip-Flop

Dual-Edge Triggering: Latch-Mux

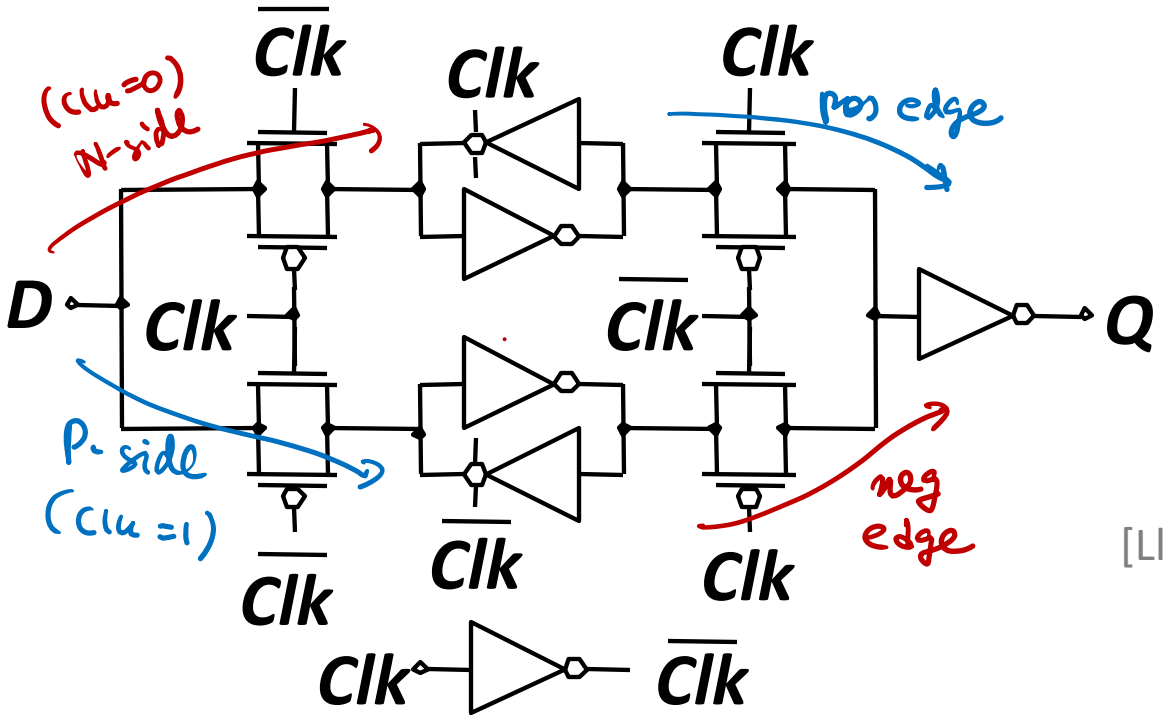
- **Saves clocking energy *regardless* of data activity!**
- **Extra Mux → longer delay**





DET Latch-Mux: Circuit Example

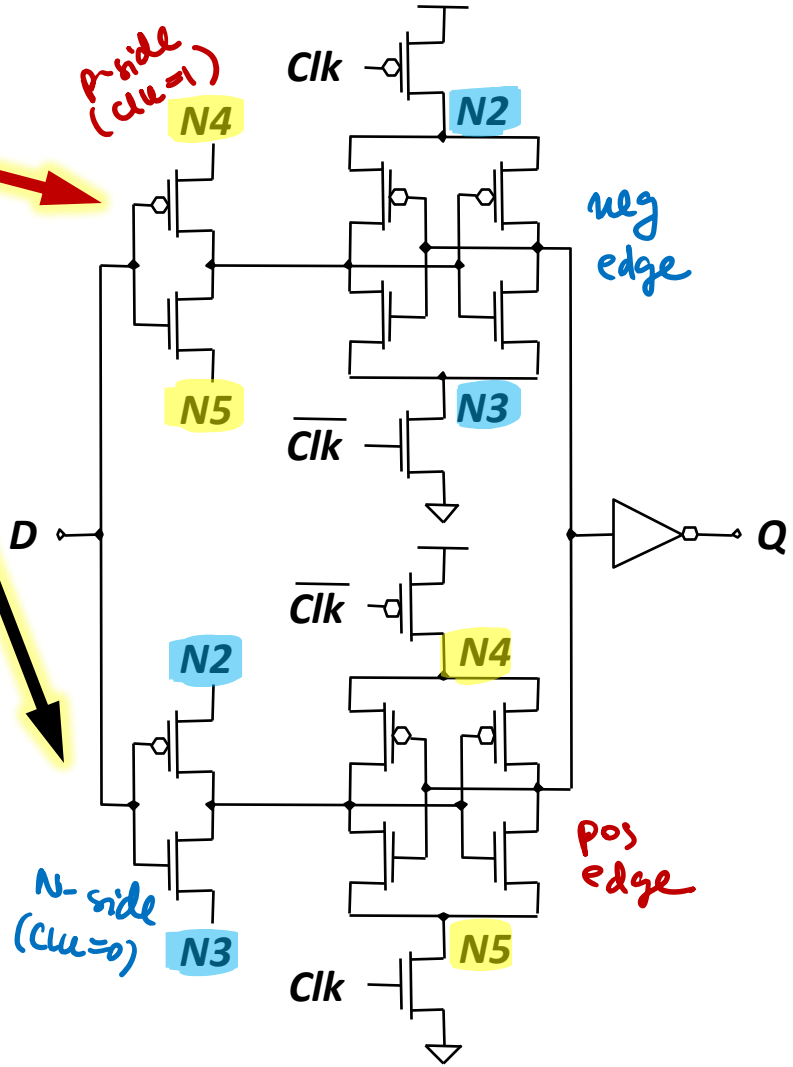
- **Pass-gate latches**
  - One transparent when *Clk* = 0
  - One transparent when *Clk* = 1
- **Pass-gate mux selects the output of the opaque latch**



[Llopis and Sachdev, 1996]

C²MOS Latch-Mux (C²MOS-LM)

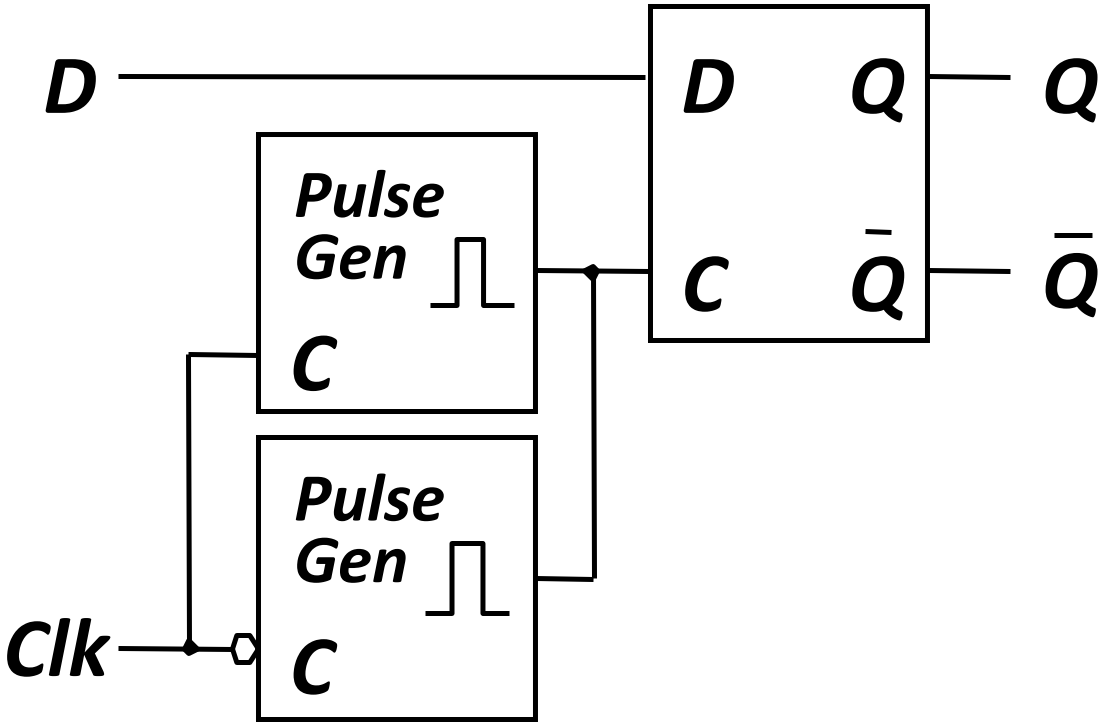
- **C²MOS latches**
  - Transparent when ***Clk* = 1**
  - Transparent when ***Clk* = 0**
- **Mux: two C²MOS inverters that propagate the output of the opaque latch**
- **Large clock transistors shared between the latches and the multiplexer**



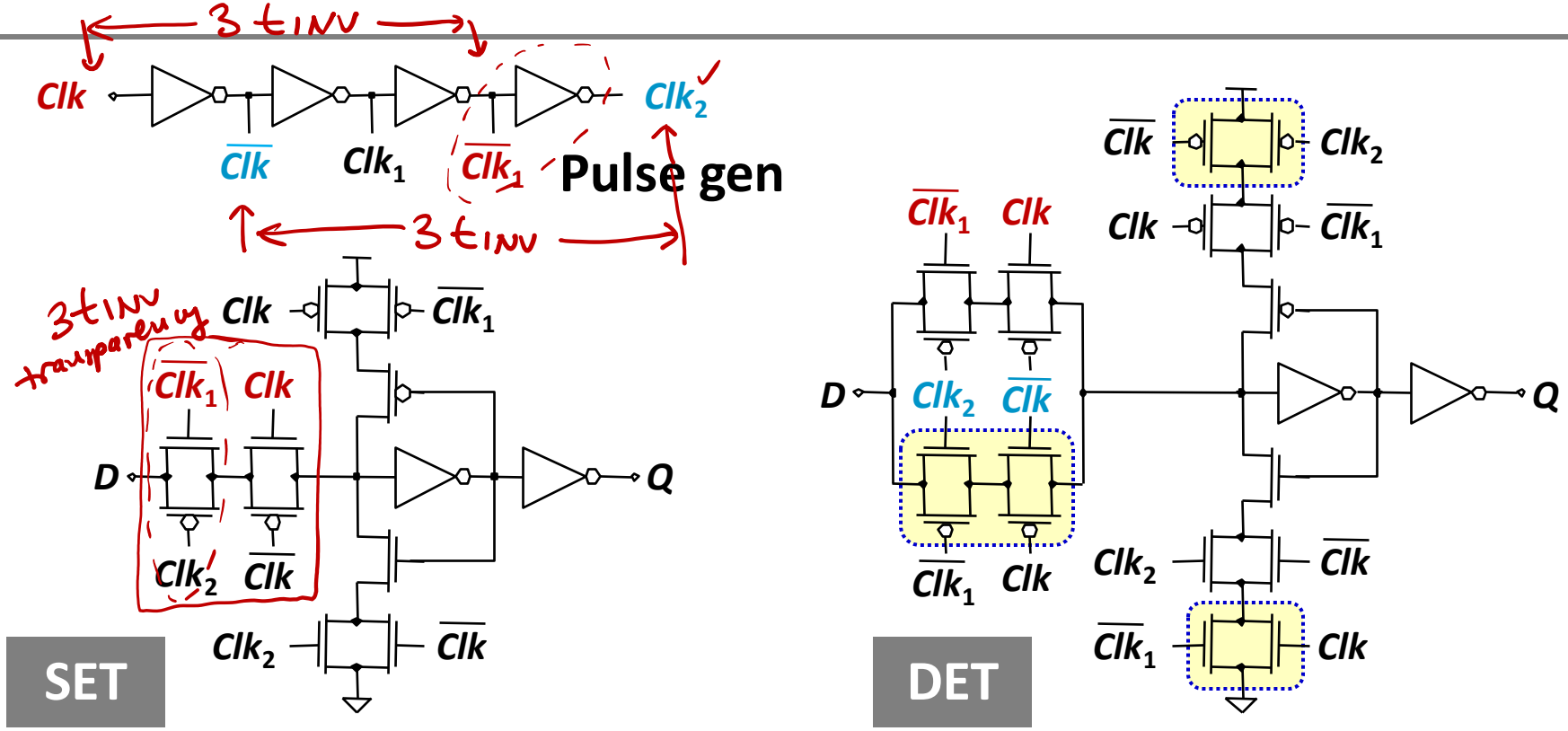
[Gago et al., 1993]

Dual-Edge Triggering: Pulsed-Latch

- First stage: pulse generator
- Second stage: capturing latch



DET Pulsed-Latch: Circuit Example



- Pulse gen transparent to *D* when $Clk = \overline{Clk_1} = 1$, or when $\overline{Clk} = Clk_2 = 1$ ⇒ shortly after both edges of the clock

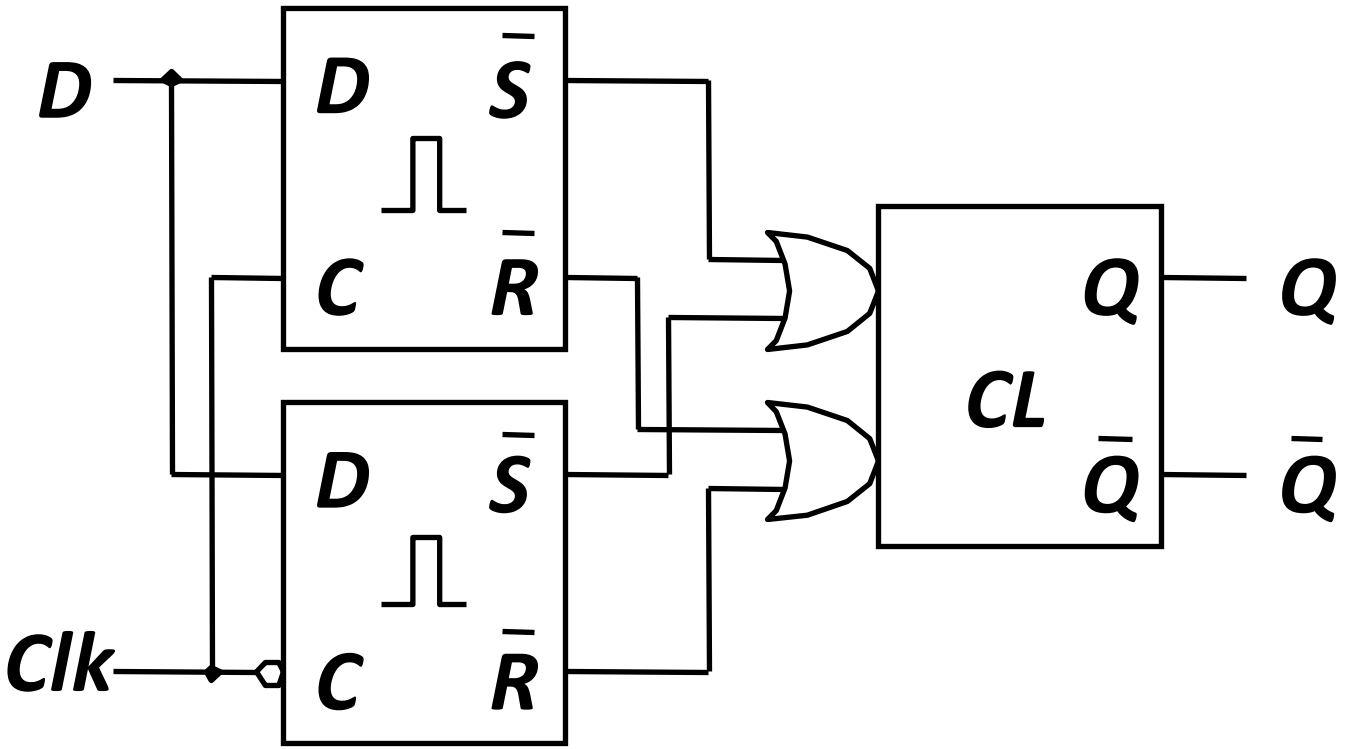
3 tINV

3 tINV

- DET PL consumes lot of energy for 4 clocked pass gates

Dual-Edge Triggered Flip-Flop

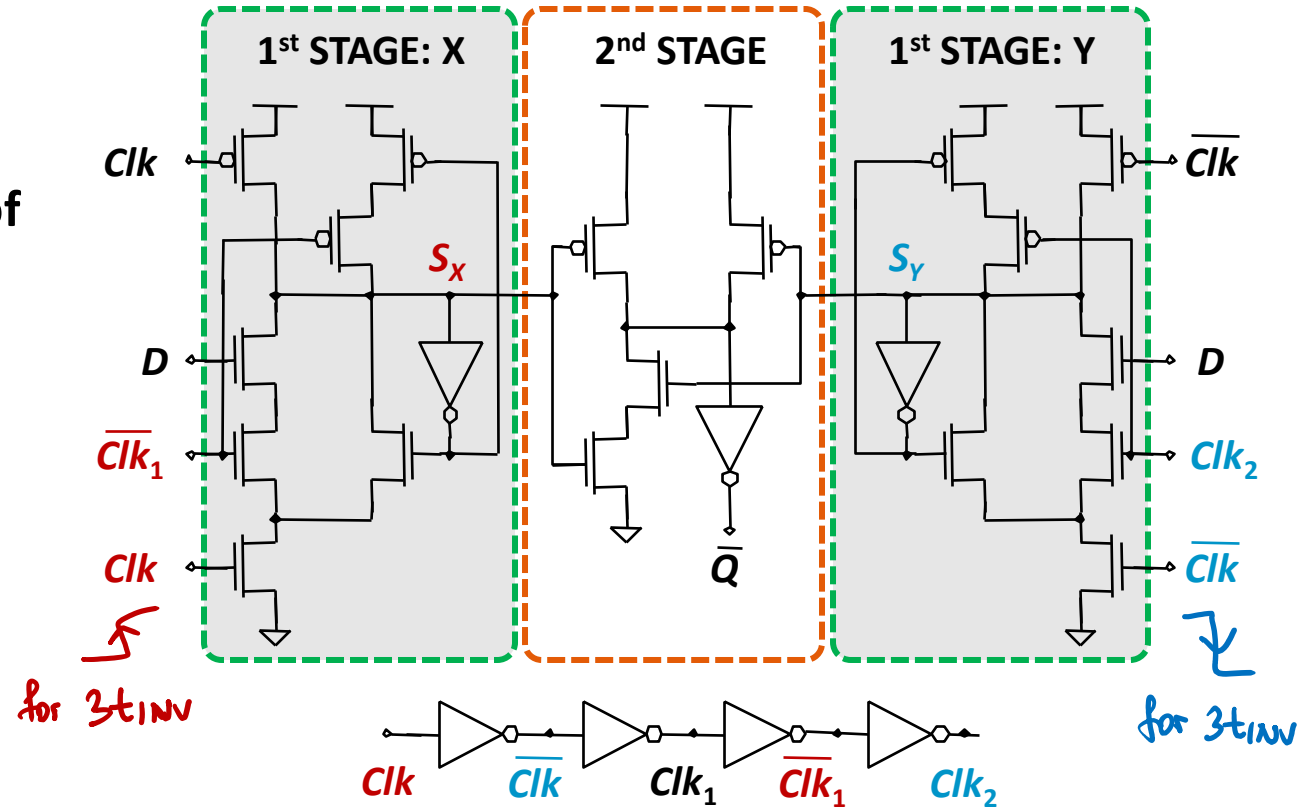
- **Based on SR latch**



DET Flip-Flop Circuit Example

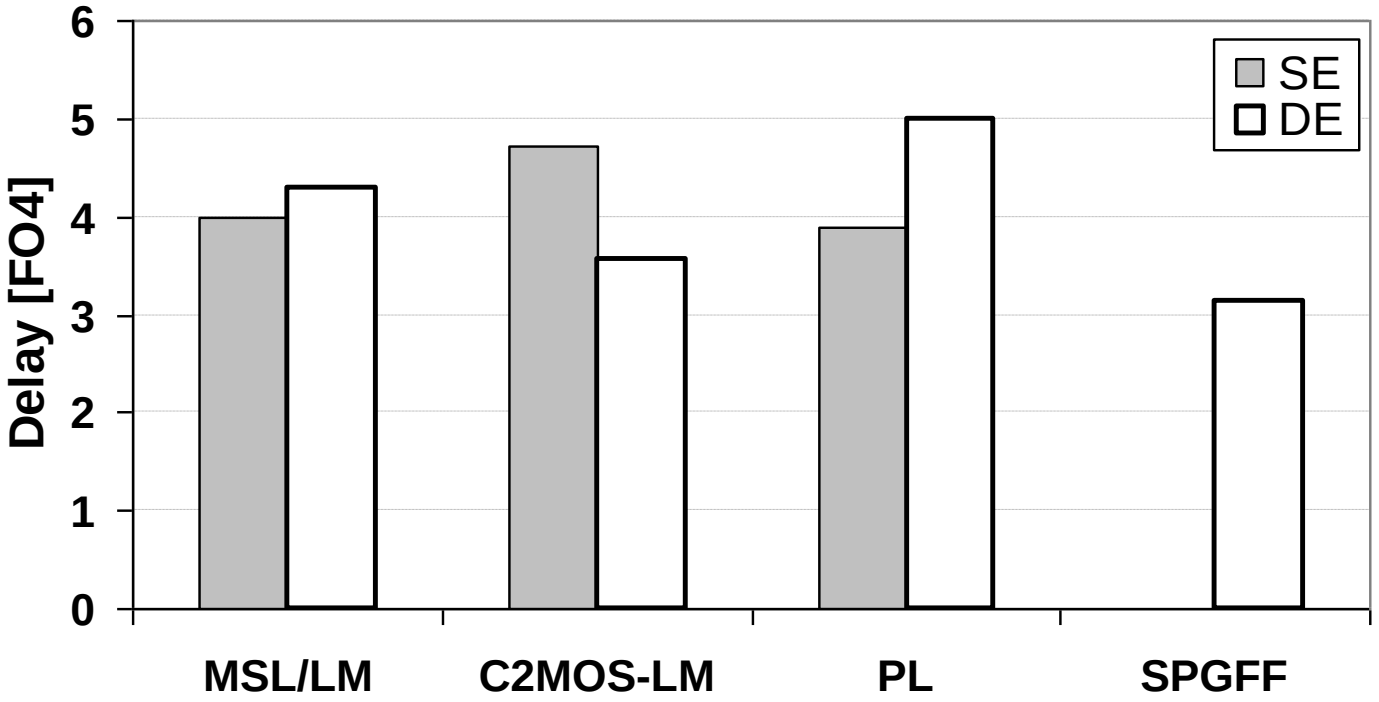
S_X and S_Y alternately precharge / evaluate

- At any moment, one of S_X and S_Y keeps the value of data sampled at the most recent clock edge
- The other S_X or S_Y is precharged high



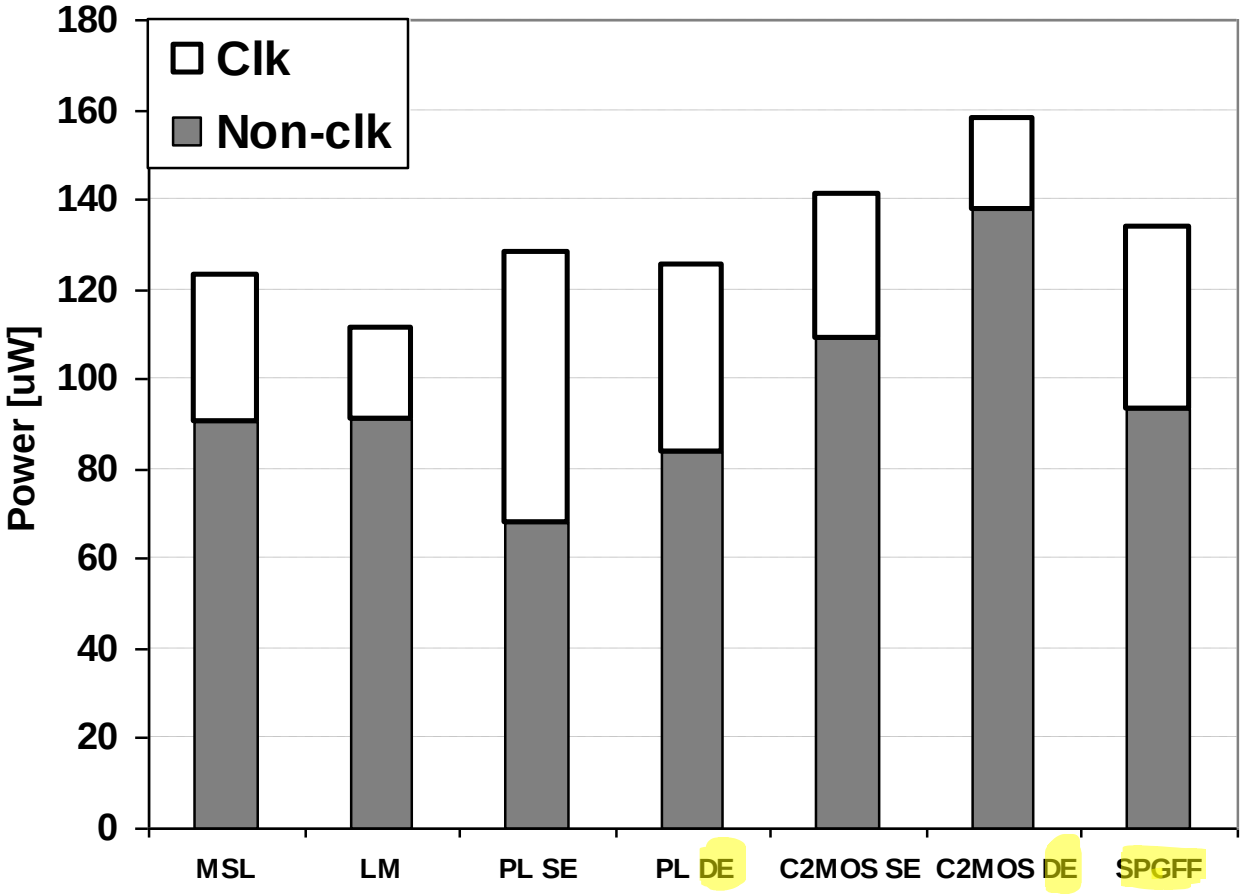
- Pulses at S_X and S_Y have same width as clock
- 2nd stage is a simple NAND gate (no need for a latch)

SET vs. DET: Delay Comparison



- LMs have two equally critical paths
- PL more complex, adding more capacitance to the critical path compared to SET PL
- SPGFF has short domino-like critical path ⇒ fastest

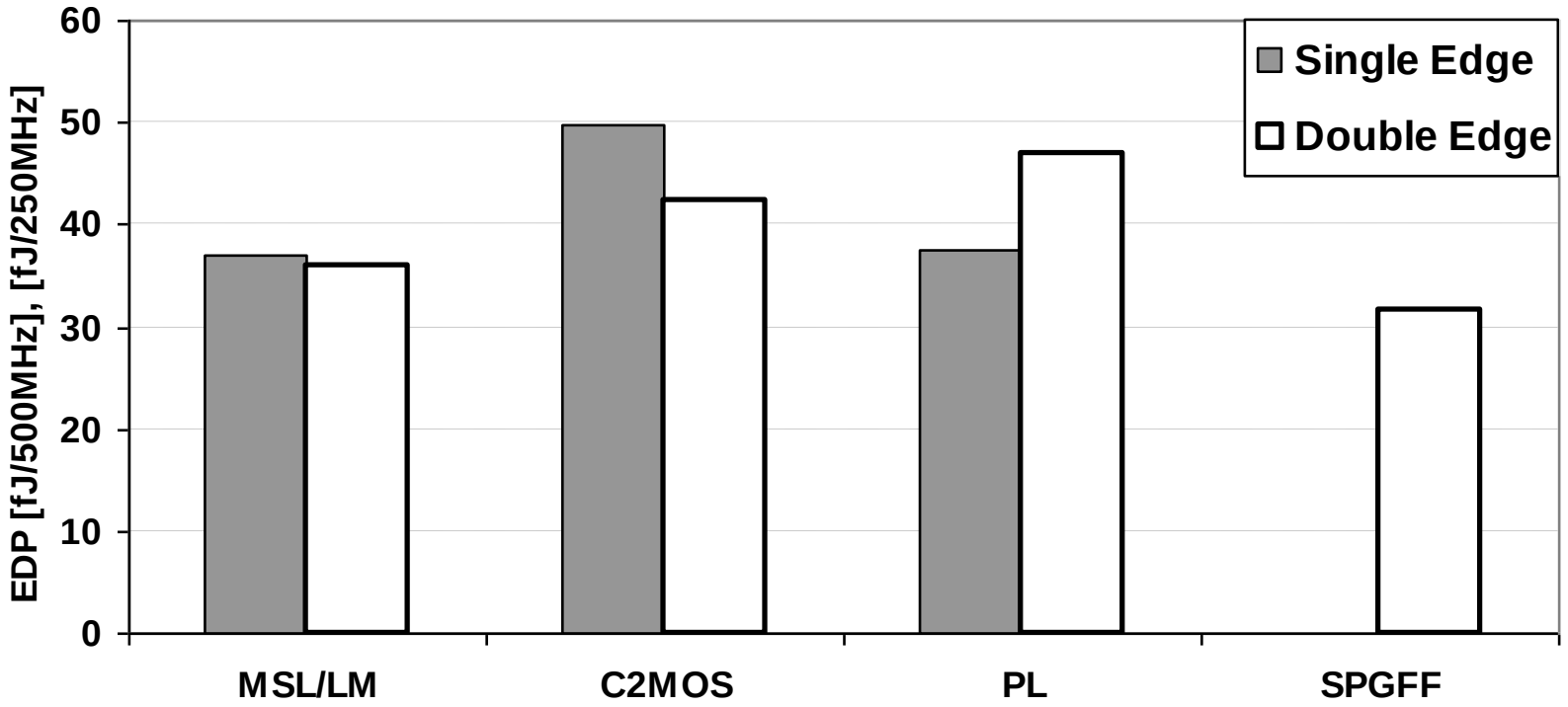
SET vs. DET: Power Consumption



0.18 µm, 500MHz for SET, 250MHZ for DET, high load

- LM's benefit from clever implementation of latch-mux structure with clock transistor sharing
- PL adds extra high-activity capacitance compared to SET PL
- SPGFF power consumption is in the middle, mainly due to alternate switching of nodes S_X and S_Y

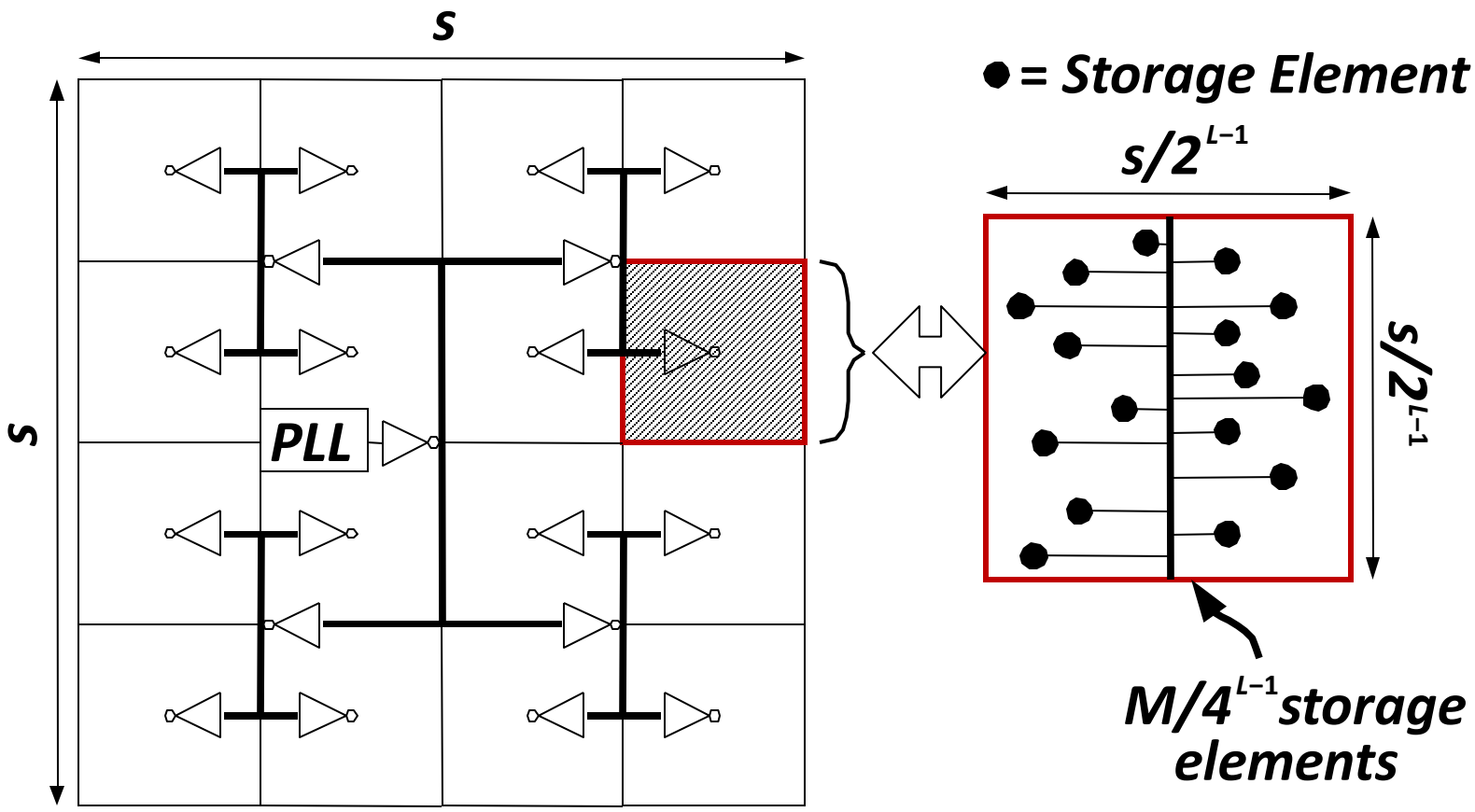
SET vs. DET: EDP Comparison



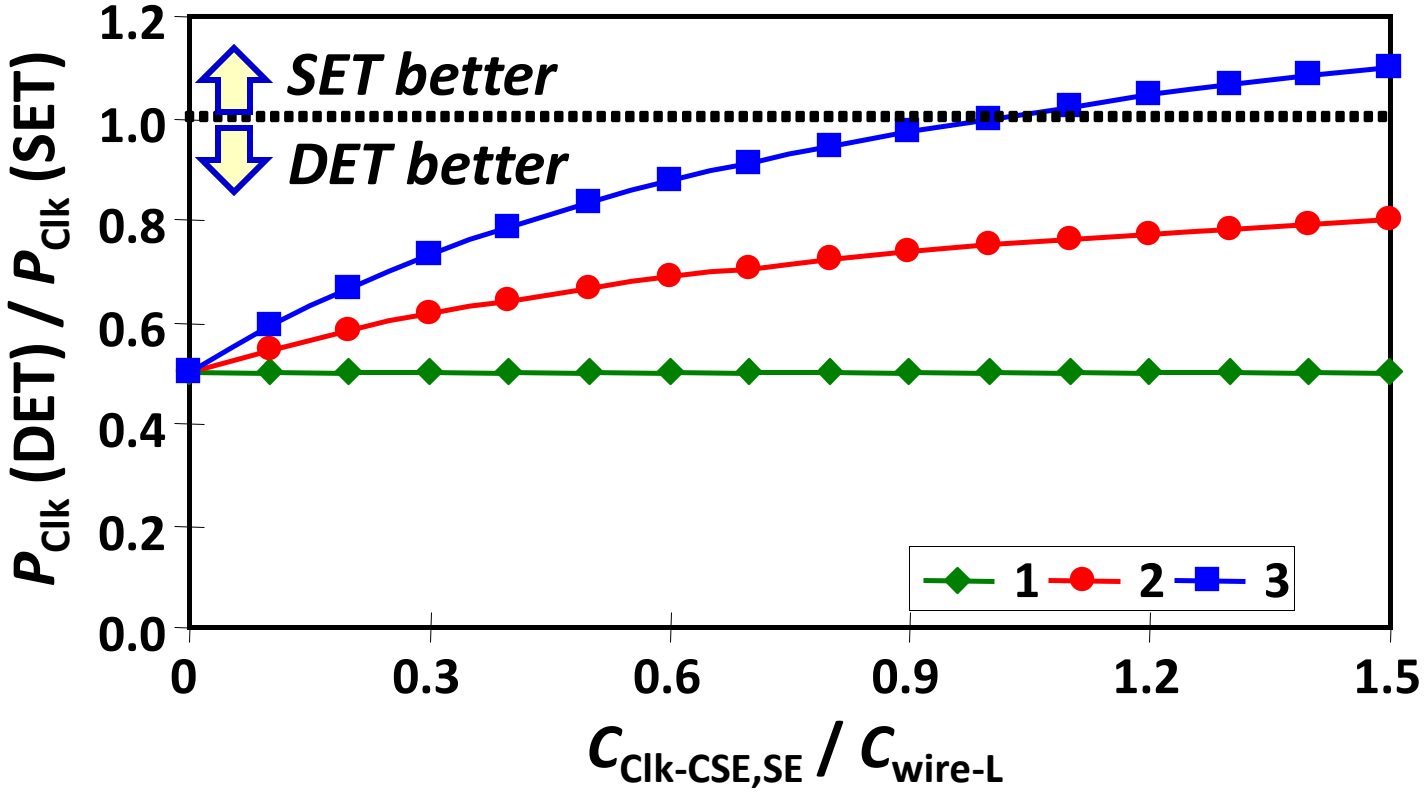
- [MED] LMs have similar or better EDP than their SET peers
- [WORST] PL hass worse delay and energy compared to SET PL, due to more complex design
- [BEST] SPGFF is fastest with moderate energy consumption: lowest EDP

Clock Distribution for DET

- **H-tree clock distribution with *L* levels of clock buffers**
  - Example below: $L = 3$



SET vs. DET: Clocking Power



DET wins if the total clock load capacitance is < 2x the capacitance of a SET design

Clocking Methodologies: Summary

- **Matched RC with active deskewing**
- **Clocking accounts for 30-50% of power**
- **Techniques for clock power reduction**
  - Global Clk gating
  - Local Clk gating
    - Data-transition look-ahead FFs
  - Low-swing clocking
  - Dual-edge triggering
- **Examine system timing and power**